

69206

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

SEARCH REQUEST FORM

Requestor's

Name:

Soubello

Serial

Number:

09/462 955

Date:

6/19

Phone:

308-6043

Art Unit:

1632

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

✓ 655-676

682-701

✓ 211-991

✓ 409-991

✓ 611-991

711-991

734-785

of seq ID

NO. 1 B

no pending

STAFF USE ONLY

Date completed:

6/20

Searcher:

S. S. L. 308-6043

Terminal time:

7

Elapsed time:

9

CPU time:

Total time:

Number of Searches:

Number of Databases:

4

Search Site

STIC

✓ CM-1

Pre-S

Type of Search

7 N.A. Sequence

A.A. Sequence

Structure

Bibliographic

Vendors

IG

STN

Dialog

APS

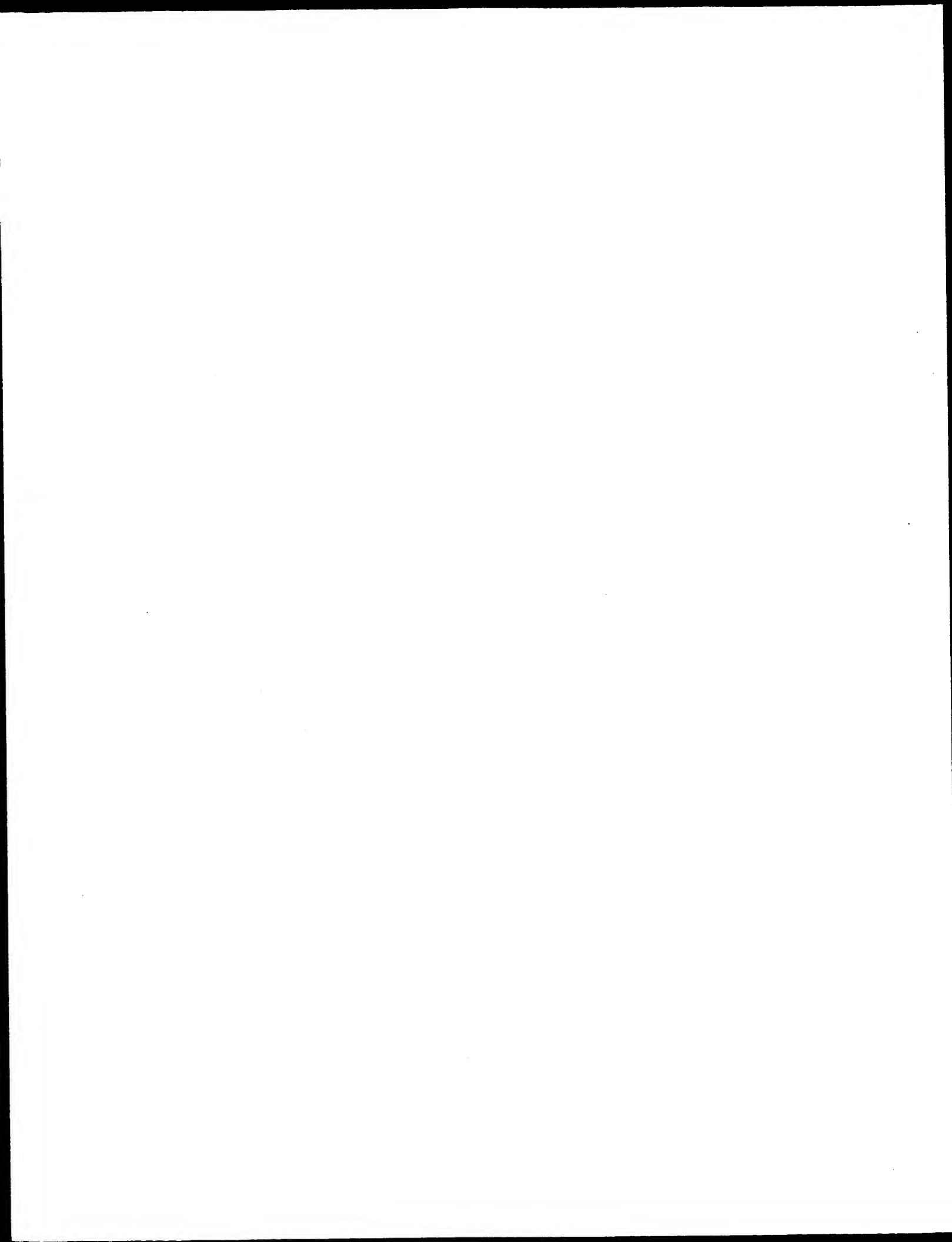
Geninfo

SDC

DARC/Questel

✓ Other

BEST AVAILABLE COPY



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:43:14 ; Search time 5904.86 Seconds
(without alignments)
70.879 Million cell updates/sec

Title: US-09-462-955B-1_COPY_682_701
Perfect score: 20
Sequence: 1 ctaccgcggccgaagccggg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_hgt_hum.*
- 31: em_hgt_inv.*
- 32: em_hgt_other.*
- 33: em_hgt_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
--------	-------	-------	-------	--------	----	----	-------------

1	20	100.0	21	6	ARI172776	ARI172776	Sequence
2	20	100.0	1291	6	ARI172770	ARI172770	Sequence
3	20	100.0	1291	14	CFDCG	CFDCG	Coconut fol
4	17.4	87.0	181	8	CISAT	CISAT	Citrus icha
5	17.4	87.0	20400	8	AF271736S1	AF271736S1	Euglena g
6	17.4	87.0	21040	8	AF271738S2	AF271738S2	Euglena g
7	17.4	87.0	21040	8	AF271738S2	AF271738S2	Euglena g
8	17	85.0	2411	10	AF2777902	AF2777902	Rattus no
9	17	85.0	3830	1	KPN9980	KPN9980	Klebsiell
10	17	85.0	31938	2	AC094656	AC094656	Rattus no
11	17	85.0	64620	2	AC103767	AC103767	Homo sapi
12	16.8	84.0	2142	1	CTV16136	CTV16136	Clostridium
13	16.8	84.0	5451	6	AX078430	AX078430	Sequence
14	16.8	84.0	6220	1	AB019552	AB019552	Aeropyrum
15	16.8	84.0	6220	1	AB078015	AB078015	Aeropyrum
16	16.8	84.0	6220	1	AB078016	AB078016	Aeropyrum
17	16.8	84.0	6575	1	AB078017	AB078017	Aeropyrum
18	16.8	84.0	6576	1	AB078022	AB078022	Aeropyrum
19	16.8	84.0	6630	1	AB078019	AB078019	Aeropyrum
20	16.8	84.0	6631	1	AB078021	AB078021	Aeropyrum
21	16.8	84.0	6632	1	AB019554	AB019554	Aeropyrum
22	16.8	84.0	6660	1	AB078020	AB078020	Aeropyrum
23	16.8	84.0	10146	1	AE004743	AE004743	Pseudomon
24	16.8	84.0	10717	1	AB008745	AB008745	Aeropyrum
25	16.8	84.0	251700	1	AF000062	AF000062	Aeropyrum
26	16.4	82.0	990	4	AY011570	AY011570	Procavia
27	16.4	82.0	990	4	AY011571	AY011571	Loxodonta
28	16	80.0	341	6	AX184404	AX184404	Sequence
29	16	80.0	5135	1	TTNUSA1NF	TTNUSA1NF	T thermophi
30	16	80.0	10305	1	AE009820	AE009820	Pyrobacul
31	16	80.0	16140	1	AE007014	AE007014	Mycobacte
32	16	80.0	20431	1	MTCY2B12	MTCY2B12	Mycobacteri
33	16	80.0	169329	2	AL592211	AL592211	Homo sapi
34	15.8	79.0	337	1	AF012803	AF012803	Saccharom
35	15.8	79.0	580	8	AF019821	AF019821	Hemarthri
36	15.8	79.0	633	1	AF004392	AF004392	Aeromonas
37	15.8	79.0	979	10	RATPAM28	RATPAM28	Rattus norv
38	15.8	79.0	1821	8	CEY16012	CEY16012	Calymperes
39	15.8	79.0	2595	1	AB029372	AB029372	Thermus t
40	15.8	79.0	2614	1	AF227236	AF227236	Halobacte
41	15.8	79.0	2929	10	RNPAM202	RNPAM202	X59686 Rat
42	15.8	79.0	3212	10	RNPAM203	RNPAM203	X59687 Rat
43	15.8	79.0	3225	6	E03428	E03428	cDNA sequen
44	15.8	79.0	3226	6	AR036184	AR036184	Sequence
45	15.8	79.0	3226	10	RNPAM205	RNPAM205	X59689 Rat

ALIGNMENTS

RESULT	1	ARI172776	Sequence	7 from patent US 6303345.	21 bp	DNA	Linear	PAT 17-DEC-2001
ARI172776	LOCUS	ARI172776	Sequence	7 from patent US 6303345.	21 bp	DNA	Linear	PAT 17-DEC-2001
DEFINITION	ARI172776	Sequence	7 from patent US 6303345.	21 bp	DNA	Linear	PAT 17-DEC-2001	
ACCESSION	ARI172776	Sequence	7 from patent US 6303345.	21 bp	DNA	Linear	PAT 17-DEC-2001	
VERSION	ARI172776.1	GI:17912267	Sequence	7 from patent US 6303345.	21 bp	DNA	Linear	PAT 17-DEC-2001
KEYWORDS	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.
SOURCE	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.
ORGANISM	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.	Unknown.
REFERENCE	1 (bases 1 to 21)	Unclassified.	Unclassified.	Unclassified.	Unclassified.	Unclassified.	Unclassified.	Unclassified.
AUTHORS	Rohde, W., Becker, D., Randles, J.W., Hehn, A. and Salamini, F.							
TITLE	Use of a virus DNA as promoter							
JOURNAL	Patent: US 6303345-A 7 16-OCT-2001;							
FEATURES	Location/Qualifiers							
source	1..21							
BASE COUNT	4 a 8 c 8 g 1 t							
ORIGIN								

Query Match 100.0%; Score 20; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccgcccgaagccggg 20
|||||
Db 1 CTACCCGCGCGAAGCCGGG 20

RESULT 2

ARI172770
LOCUS ARI172770 1291 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 1 from patent US 6303345.
ACCESSION ARI172770
VERSION ARI172770.1 GI:17912261
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1291)
AUTHORS Rohde, W., Becker, D., Randles, J. W., Hehn, A. and Salamini, F.
TITLE Use of a virus DNA as promoter
JOURNAL Patent: US 6303345-A 1 16-OCT-2001;
FEATURES
1. Location/Qualifiers
1. .1291
/organism="unknown"
BASE COUNT 336 a 323 c 332 g 300 t
ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 1291;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccgcccgaagccggg 20
|||||
Db 1072 CTACCCGCGCGAAGCCGGG 1091

RESULT 3

CFDCG
LOCUS CFDCG 1291 bp ss-DNA circular VRL 02-AUG-1993
DEFINITION Coconut foliar decay virus, complete genome.
ACCESSION M29963
VERSION M29963.1 GI:323306
KEYWORDS circular; complete genome.
SOURCE Coconut foliar decay virus
ORGANISM Coconut foliar decay virus
VIRUSES ssDNA viruses; Nanovirus.

REFERENCE 1 (bases 1 to 1291)
AUTHORS Rohde, W., Randles, J. W., Langridge, P. and Hanold, D.
TITLE Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus
JOURNAL Virology 176, 648-651 (1990)
MEDLINE 90266484
COMMENT Draft entry and printed sequence for [1] kindly submitted by W. Rhode, 15-MAR-1989, for release after publication.

FEATURES
source
1. .1291
/organism="Coconut foliar decay virus"
/db_xref="taxon:12474"
40-70
/note="stem-loop structure"
stem_loop 103. .975
CDS
/note="ORF 1"
/codon_start=1
/protein_id="AAA42894.1"
/db_xref="GI:323307"
/translation="MGSIRWCFGLNVEEAAVVRRESLNLVVAIVGDEVAPS
TGQRHQFHLKTRRLOGLKTVLGNDRHLETRGSDQNRDYSKERVLLHGV
TPGVKPRLAORFAEEDELRLEDGGYRCVVGASVEMRWAAENPFFPHNQ
LEVAISAGEADRTILWICRGDGGKSVFAYLGLKPDWFCGTGTRKDLVLYE
DPKRLILDPVRCRLELYNLALLECVKNRAFSDDKYBPLSYLGLGDHVLVFNVLDP
YLKTSRDRILKWN1"
314. .775

CDS

/note="ORF 2"
/codon_start=1
/protein_id="AAA42895.1"
/db_xref="GI:323308"
/translation="MTGFTWSRPVPTNRLETTVRNGCFSTTESRLVLESKGDWPN
DLLRLNMSAWKQADTEDALYTELRWNGQDGLKIRSHFHIIGSLKCCLRSESQRT
IAQSGYADETEETGSPCLNPISDSPTGTHVVEPTCYTSTSTQNEI"
complement(422. .568)
/note="ORF 6"
/codon_start=1
/protein_id="AAA42896.1"
/db_xref="GI:323309"
/translation="MEMGTDFQRPILSPKPLRVORIGIRLGLPGGVHQPQOIVGP
IVAF"
639. .797
/note="ORF 3"
/codon_start=1
/protein_id="AAA42897.1"
/db_xref="GI:323310"
/translation="MTRRRRRREVRCVQISRTQARLVLHMMWNOKGRIVPVHRGPKTK
FNPRCTOV"
complement(823. .987)
/note="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="MTHLTNIPQFYSVPADFQIIRQDYGKYEYMMVEPKITKGFVFP
RTECPVLNTF"
1098. .1286
/note="ORF 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRVMGGPTINDSIWIRTNLLCLQCQPLSTSPQVSSLEKK
AASLVPSICFCAIGRLS"
BASE COUNT 336 a 323 c 332 g 300 t
ORIGIN

Query Match 100.0%; Score 20; DB 14; Length 1291;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccgcccgaagccggg 20
|||||
Db 1072 CTACCCGCGCGAAGCCGGG 1091

RESULT 4

CISAT
LOCUS CISAT 181 bp DNA linear PLN 29-OCT-1996
DEFINITION Citrus ichangensis satellite DNA.
ACCESSION X51946
VERSION X51946.1 GI:1653975
KEYWORDS extrachromosomal DNA; satellite DNA.
SOURCE Citrus ichangensis.
ORGANISM Citrus ichangensis.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Sapindales; Rutaceae; Citrus.
Beridze, T.G.
REFERENCE 1 (bases 1 to 181)
AUTHORS Beridze, T.G.
TITLE Direct Submission
JOURNAL Submitted (21-PEB-1990) Beridze T.G., Institute of Plant
Biochemistry, Georgian Academy of Science, Tbilisi 380031, USSR
REMARK Revised by author 29-OCT-96
2 (bases 1 to 181)
AUTHORS Beridze, T., Tsirekidze, N. and Roytberg, M.A.
TITLE On the tertiary structure of satellite DNA
JOURNAL Blochimie 74 (2), 187-194 (1992)
MEDLINE 92256540
COMMENT On Oct 30, 1996 this sequence version replaced gi:18037.
FEATURES Location/Qualifiers

source
1. .181
/organism="Citrus ichangensis"
/db_xref="taxon:2709"
/tissue_type="leaves"

misc_feature
40 a 54 c 64 g 23 t
/note="satellite DNA"

BASE COUNT 40 a 54 c 64 g 23 t
ORIGIN
Query Match 87.0%; Score 17.4; DB 8; Length 181;
Best Local Similarity 94.7%; Pred. No. 7.6e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 taccggcgccgaagcgccgg 20
||||| |||||||
Db 26 TACCGCGCCGAAGCGCGG 44

RESULT 5
AF271736S1 20400 bp DNA linear PLN 02-MAY-2001
LOCUS
DEFINITION
Euglena gracilis gamma-tubulin gene, exons 1-14 sequence.
ACCESSION
AF271736
VERSION
AF271736.1 GI:13661562
KEYWORDS
1 of 2
SEGMENT
Euglena gracilis.
SOURCE
Euglena gracilis.
ORGANISM
Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
REFERENCE
1 (bases 1 to 20400)
AUTHORS
Canaday,J., Tessier,L.H., Imbault,P. and Paulus,F.
TITLE
Analysis of Euglena gracilis alpha-, beta- and gamma-tubulin genes:
introns and pre-mRNA maturation
Mol. Genet. Genomics 265 (1), 153-160 (2001)
JOURNAL
MEDLINE
21263010
REFERENCE
2 (bases 1 to 20400)
AUTHORS
Paulus,F.P. and Tessier,L.H.
TITLE
Direct Submission
JOURNAL
Submitted (23-MAY-2000) IBMP, CNRS, 12 rue du General Zimmer,
Strasbourg 67084, France

FEATURES
source
1. .20400
Location/Qualifiers
/organism="Euglena gracilis"
/strain="Klebs Z"
/db_xref="taxon:3039"
814. .869
/number=1
exon
8142. .8254
/number=2
exon
9340. .9393
/number=3
exon
10800. .10913
/number=4
exon
11103. .11171
/number=5
exon
12490. .12569
/number=6
exon
13657. .13702
/number=7
exon
14064. .14150
/number=8
exon
14705. .14791
/number=9
exon
17219. .17368
/number=10
exon
17477. .17548
/number=11
exon
17782. .17856
/number=12
exon
18317. .18384
/number=13
exon
19559. .19652

BASE COUNT 5172 a 4729 c 4709 g 5790 t
ORIGIN
/number=14

Query Match 87.0%; Score 17.4; DB 8; Length 20400;
Best Local Similarity 94.7%; Pred. No. 1.6e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 taccggcgccgaagcgccgg 20
||||| |||||||
Db 11876 TGCCGCGCGAAGCGCGG 11894

RESULT 6
AF271738S2 21040 bp DNA linear PLN 23-APR-2001
LOCUS
DEFINITION
Euglena gracilis gamma-tubulin gene, exons 2-17, complete cds.
ACCESSION
AF271739
VERSION
AF271739.1 GI:13661567
KEYWORDS
2 of 2
SEGMENT
Euglena gracilis.
SOURCE
Euglena gracilis.
ORGANISM
Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
REFERENCE
1 (bases 1 to 21040)
AUTHORS
Canaday,J., Tessier,L.H., Imbault,P. and Paulus,F.
TITLE
Analysis of Euglena gracilis alpha-, beta- and gamma-tubulin genes:
introns and pre-mRNA maturation
Mol. Genet. Genomics 265 (1), 153-160 (2001)
JOURNAL
MEDLINE
21263010
REFERENCE
2 (bases 1 to 21040)
AUTHORS
Paulus,F.P. and Tessier,L.H.
TITLE
Direct Submission
JOURNAL
Submitted (24-MAY-2000) IBMP, CNRS, 12 rue du General Zimmer,
Strasbourg 67084, France

FEATURES
source
1. .21040
Location/Qualifiers
/organism="Euglena gracilis"
/strain="Klebs Z"
/db_xref="taxon:3039"
join(AF271738.1:1576. .1631,5406. .5518,6546. .6599,
8036. .8149,8339. .8407,9728. .9807,10965. .11010,
11366. .11452,12006. .12092,14512. .14661,14769. .14840,
15074. .15148,15563. .15630,16802. .16895,17682. .17726,
19824. .19894,20826. .21040)
/product="gamma-tubulin"
join(AF271738.1:1583. .1631,5406. .5518,6546. .6599,
8036. .8149,8339. .8407,9728. .9807,10965. .11010,
11366. .11452,12006. .12092,14512. .14661,14769. .14840,
15074. .15148,15563. .15630,16802. .16895,17682. .17726,
19824. .19894,20826. .20937)
/codon_start=1
/product="gamma-tubulin"
/protein_id="AAK37967.1"
/db_xref="GI:13661568"

translation="MPREITLTCGCGGNGVTFWKLCAEHGIPDGLIODYATILG
GDRKDYFYQADDEHYIPRALLDLEPRVINTRTGEYRNLYNPENYIUSEDGGGAGN
NWAVEKADKSDSDIVMDLREADSGESLEGVLVHSTAGTGGSGSYLLEKLNDR
YPKKLVQTSVFPNDSKESGVVQVNSLLTKRLTLNADCVVLDNTALNRYVE
RLHPSPFTSQTNALSTVMAASTTTLRYPGYNNDLIGLIASLIPTFTRCHFLMTGYT
PI*IDPNIOMTVRKTSVLDVMRLLSPKNIMVSASTRRGCYVSILNIQGVDPGOVH
KSLQIRERKMATFIPWGPASIOVALSRKSPYLETAHRYVSGMLNHTSIANI.FARTL
SNYDKLRKRSALFDVYTKIDLFKDXGLEEDSSREIVNLYVEYKACERPDYINWGSSM
YNAEGGTAPMM"
5406. .5518
/number=2
exon
6546. .6599
/number=3
exon
8036. .8149
/number=4
exon
8339. .8407
/number=5

On Dec 20, 2001 this sequence version replaced gi:17062140.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GBCW
 Center clone name: CH230-5B20
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to
 findPhrapList
 Consensus quality: 26514 bases at least Q40
 Consensus quality: 29012 bases at least Q30
 Consensus quality: 31138 bases at least Q20
 Estimated insert size: 15885; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-gel estimation
 Quality coverage: 0.2x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 20 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 1 1998: contig of 1998 bp in length
 2098: gap of unknown length
 2099: contig of 2525 bp in length
 4623: gap of unknown length
 4624: contig of 2122 bp in length
 4724: gap of unknown length
 5943: gap of unknown length
 6846: contig of 2123 bp in length
 6946: gap of unknown length
 9069: contig of 1766 bp in length
 9169: contig of 1622 bp in length
 10934: gap of unknown length
 11034: contig of 1599 bp in length
 12657: contig of 1599 bp in length
 12757: gap of unknown length
 14356: contig of 1105 bp in length
 14456: gap of unknown length
 15561: contig of 1360 bp in length
 15661: contig of 1360 bp in length
 17021: contig of 1164 bp in length
 17121: gap of unknown length
 18285: contig of 1270 bp in length
 18385: gap of unknown length
 19755: contig of 1511 bp in length
 19755: gap of unknown length
 21265: contig of 1427 bp in length
 21266: gap of unknown length
 21366: contig of 1742 bp in length
 22793: gap of unknown length
 22893: contig of 1742 bp in length
 24635: gap of unknown length
 24735: contig of 1060 bp in length
 25794: gap of unknown length
 25795: contig of 1042 bp in length
 25895: gap of unknown length
 26937: contig of 1438 bp in length
 27037: gap of unknown length
 28474: contig of 1137 bp in length
 28575: gap of unknown length
 29711: contig of 1001 bp in length
 29712: gap of unknown length
 30812: contig of 1026 bp in length.
 30813: Location/Qualifiers
 30913 1..31938
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-5B20"

BASE COUNT 7910 a 6842 c 7612 g 7622 t 1952 others
 ORIGIN
 Query Match 85.0%; Score 17; DB 2; Length 31938;
 Best Local Similarity 100.0%; Pred. No. 2e+03;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 cccggccgaagccggg 20
 |||||
 Db 24061 CCCGCCGAGCGCGG 24077
 |||||
 RESULT 10
 AC103767
 LOCUS
 DEFINITION Homo sapiens chromosome 8 clone RP11-1077M7 map 8, LOW-PASS
 SEQUENCE SAMPLING.
 AC103767
 AC103767.1 GI:17149652
 HTG: HTGS_PHASE0.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 64620)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 8, clone RP11-1077M7
 UNPUBLISHED
 2 (bases 1 to 64620)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Jones,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
 McNeil,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L21727
 Center clone name: 1077_M_7

 * NOTE: This record contains 83 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows

FEATURES
 source

* overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 675: contig of 675 bp in length
 676 775: gap of 100 bp
 776 1433: contig of 658 bp in length
 1434 1533: gap of 100 bp
 1534 2238: contig of 705 bp in length
 2239 2338: gap of 100 bp
 2339 3016: contig of 678 bp in length
 3017 3116: gap of 100 bp
 3117 3772: contig of 656 bp in length
 3773 3872: gap of 100 bp
 3873 4508: contig of 636 bp in length
 4509 4608: gap of 100 bp
 4609 5270: contig of 662 bp in length
 5271 5370: gap of 100 bp
 5371 6038: contig of 668 bp in length
 6039 6138: gap of 100 bp
 6139 6837: contig of 699 bp in length
 6838 6937: gap of 100 bp
 6938 7613: contig of 676 bp in length
 7614 7713: gap of 100 bp
 7714 8402: contig of 689 bp in length
 8403 8502: gap of 100 bp
 8503 9179: contig of 677 bp in length
 9180 9279: gap of 100 bp
 9280 9942: contig of 663 bp in length
 9943 10042: gap of 100 bp
 10043 10647: contig of 605 bp in length
 10648 10747: gap of 100 bp
 10748 11391: contig of 644 bp in length
 11392 11491: gap of 100 bp
 11492 12174: contig of 683 bp in length
 12175 12274: gap of 100 bp
 12275 12967: contig of 693 bp in length
 12968 13067: gap of 100 bp
 13068 13749: contig of 682 bp in length
 13750 13849: gap of 100 bp
 13850 14558: contig of 709 bp in length
 14559 14658: gap of 100 bp
 14659 15347: contig of 689 bp in length
 15348 15447: gap of 100 bp
 15448 16153: contig of 706 bp in length
 16154 16253: gap of 100 bp
 16254 16948: contig of 695 bp in length
 16949 17048: gap of 100 bp
 17049 17740: contig of 692 bp in length
 17741 17840: gap of 100 bp
 17841 18523: contig of 683 bp in length
 18524 18623: gap of 100 bp
 18624 19286: contig of 663 bp in length
 19287 19386: gap of 100 bp
 19387 20058: contig of 672 bp in length
 20059 20158: gap of 100 bp
 20159 20821: contig of 663 bp in length
 20822 20921: gap of 100 bp
 20922 21619: contig of 698 bp in length
 21620 21719: gap of 100 bp
 21720 22423: contig of 704 bp in length
 22424 22523: gap of 100 bp
 22524 23226: contig of 703 bp in length
 23227 23326: gap of 100 bp
 23327 24014: contig of 688 bp in length
 24015 24114: gap of 100 bp
 24115 24806: contig of 692 bp in length
 24807 24906: gap of 100 bp
 24907 25598: contig of 692 bp in length
 25599 25698: gap of 100 bp
 25699 26385: contig of 687 bp in length
 26386 26485: gap of 100 bp

* 26486 27145: contig of 660 bp in length
 * 27146 27245: gap of 100 bp
 * 27246 27936: contig of 691 bp in length
 * 27937 28036: gap of 100 bp
 * 28037 28719: contig of 683 bp in length
 * 28720 28819: gap of 100 bp
 * 28820 29475: contig of 656 bp in length
 * 29476 29575: gap of 100 bp
 * 29576 30265: contig of 690 bp in length
 * 30266 30365: gap of 100 bp
 * 30366 31064: contig of 699 bp in length
 * 31065 31164: gap of 100 bp
 * 31165 31862: contig of 698 bp in length
 * 31863 31962: gap of 100 bp
 * 31963 32676: contig of 714 bp in length
 * 32677 32776: gap of 100 bp
 * 32777 33451: contig of 675 bp in length
 * 33452 33551: gap of 100 bp
 * 33552 34244: contig of 693 bp in length
 * 34245 34344: gap of 100 bp
 * 34345 35035: contig of 691 bp in length
 * 35036 35135: gap of 100 bp
 * 35136 35798: contig of 663 bp in length
 * 35799 35898: gap of 100 bp
 * 35899 36566: contig of 668 bp in length
 * 36567 36666: gap of 100 bp
 * 36667 37335: contig of 669 bp in length
 * 37336 37435: gap of 100 bp
 * 37436 38084: contig of 649 bp in length
 * 38085 38184: gap of 100 bp
 * 38185 38877: contig of 693 bp in length
 * 38878 38977: gap of 100 bp
 * 38978 39675: contig of 698 bp in length
 * 39676 39775: gap of 100 bp
 * 39776 40428: contig of 653 bp in length
 * 40429 40528: gap of 100 bp
 * 40529 41202: contig of 674 bp in length
 * 41203 41302: gap of 100 bp
 * 41303 41991: contig of 689 bp in length
 * 41992 42091: gap of 100 bp
 * 42092 42760: contig of 669 bp in length
 * 42761 42860: gap of 100 bp
 * 42861 43518: contig of 658 bp in length
 * 43519 43618: gap of 100 bp
 * 43619 44309: contig of 691 bp in length
 * 44310 44409: gap of 100 bp
 * 44410 45082: contig of 673 bp in length
 * 45083 45182: gap of 100 bp
 * 45183 45862: contig of 680 bp in length
 * 45863 45962: gap of 100 bp
 * 45963 46613: contig of 651 bp in length
 * 46614 46713: gap of 100 bp
 * 46714 47378: contig of 665 bp in length
 * 47379 47478: gap of 100 bp
 * 47479 48175: contig of 697 bp in length
 * 48176 48275: gap of 100 bp
 * 48276 48935: contig of 660 bp in length
 * 48936 49035: gap of 100 bp
 * 49036 49744: contig of 709 bp in length
 * 49745 49844: gap of 100 bp
 * 49845 50534: contig of 690 bp in length
 * 50535 50634: gap of 100 bp
 * 50635 51332: contig of 698 bp in length
 * 51333 51432: gap of 100 bp
 * 51433 52111: contig of 679 bp in length
 * 52112 52211: gap of 100 bp
 * 52212 52872: contig of 661 bp in length
 * 52873 52972: gap of 100 bp
 * 52973 53605: contig of 633 bp in length

Query Match 85.0%; Score 17; DB 2; Length 64620;
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ctaccggcgccgaaggcc 17
    |||||||
Db 4449 CTACCGCGCGGAGGCC 4465

RESULT 11
LOCUS CTY16136 2142 bp DNA linear BCT 23-FEB-2001
DEFINITION Clostridium thermoaceticum enr gene.
ACCESSION Y16136
VERSION Y16136.1 GI:3402835
KEYWORDS 2-enoate reductase; enr gene.
SOURCE Moorella thermoacetica.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group;
Thermoanaerobacter group; Moorella group; Moorella.
REFERENCE 1 (bases 1 to 2142)
AUTHORS Rohdich, F., Wiese, A., Feicht, R., Simon, H. and Bacher, A.
TITLE Enolate reductases of Clostridia. CLONING, SEQUENCING, AND
EXPRESSION
JOURNAL J. Biol. Chem. 276 (8), 5779-5787 (2001)
PUBMED 11060310
REFERENCE 2 (bases 1 to 2142)
AUTHORS Rohdich, F.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-1998) F. Rohdich, TU Muenchen, Institut fuer
Organische Chemie und Biochemie, Lichtenbergstr. 4, D- 85748
Garching, FRG
REMARK revised by [3]
REFERENCE 3 (bases 1 to 2142)
AUTHORS Rohdich, F.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1998) F. Rohdich, TU Muenchen, Institut fuer
Organische Chemie und Biochemie, Lichtenbergstr. 4, D- 85748
Garching, FRG
COMMENT On Aug 7, 1998 this sequence version replaced gi:2791351.
Related location: Y09960.
FEATURES
    source
    1..2142
    /organism="Moorella thermoacetica"
    /strain="DSM 1974"
    /db_xref="taxon:1525"
    32..37
    /gene="enr"
    32..2050
    /gene="enr"
    47..2050
    /gene="enr"
    /EC_number="1.3.1.31"
    /codon_start=1
    /transl_table=11
    /evidence=experimental
    /product="2-enoate reductase"
    /protein_id="CAA76082.1"
    /db_xref="GI:3402836"
    /db_xref="SPTREMBL:O52935"
    /translation="MVAVYTLFPIKIGKVEIKNKIAMTPMVGVLGATHDGFCKRVV
DYYVERAKGTGLLITSVTKVDNEIERFKAGVPVATANPLHFIATAGELTERVHAYG
KTFILQMGFGFVAAPIILLESQVAPSPALPNFWDPSITCRELTAEVETLVQASEA
AEIAVEAGFDGVEIHAMHEGYLLDQFTIALFNRRGDKYGGALDEDTLFFPIEIVRAIKD
RVGKDFPVYLFKSNYIKDWRQGLFGENFQEKGRDVEEPFLAAKILEGAYDGEDA
DAGSYDAWYAHAPVYQKHGQCYLPLTQRLKEVVKVPVIVAGRLIEPELAEEALVKQQA
DMIAIGRLGLTDPYVWKNKVTMTGRSKNIRPGICHDGCLGRGLRPLSCTVNPACGRE
FEYAIDRAPEAKQVMYVIGGVAGVAAARVPALRGHVSLEYKSDRLGSHGVVAAVPDF
KADGRLLKWNKTELGELOVEIHLNOEVTPEFVEEKNPDVVVATGSTPAIPDIPGVN
KDKVTYVSDLLGLKRGAGDRVVIIGGLVGCETALWLAQGGKDYTIIEILDLMRAGI
PVPYNNRMRLDLKMGVKKWTSTSVLEVTDDGVTLLIGNYQKRSPLPDTVILAVGF
GADQRLYNALRDKIPNLYLIGDSREPRNLAGIWEYGVGKGI"
    misc_binding
    1212..1278
    /gene="enr"
    /bound_moiety="FAD"
    1590..1626
    misc_binding

```

```

/gene="enr"
/bound_moiety="NAD"

```

```

BASE COUNT 470 a 569 c 687 g 416 t
ORIGIN

```

```

Query Match 84.0%; Score 16.8; DB 1: Length 2142;
Best Local Similarity 90.0%; Pred. No. 5.7e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ctaccggcgccgaaggccggg 20
    |||||||
Db 1223 CTCCCGCGCGGAGGCCGGG 1204

```

RESULT 12

```

LOCUS AX078430 5451 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 5 from Patent WO0107589.
ACCESSION AX078430
VERSION AX078430.1 GI:13158086
KEYWORDS Pseudomonas sp.
SOURCE Pseudomonas sp.
ORGANISM Bacteria; Proteobacteria.
REFERENCE 1 (bases 1 to 5451)
AUTHORS Fraser, C.M., Tuemmler, B., Hoheisel, J., Duesterhoeft, A., Hilbert, H.,
Timmis, K.N., Moore, E., Straetz, M., Heim, S. and Golyshin, P.
TITLE Dna sequences encoding lipopeptide synthases
JOURNAL Patent: WO 0107589-A 5 01-FEB-2001;
THE INSTITUTE FOR GENOMIC RESEARCH (US); OIAGEN GmbH (DE);
Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE);
Deutsches Krebsforschungszentrum Stiftung des Oeffentlichen Rechts
(DE); Medizinische Hochschule Hannover (DE)

```

```

FEATURES
    source
    1..5451
    /organism="Pseudomonas sp."
    /db_xref="taxon:306"

```

```

BASE COUNT 937 a 1766 c 1807 g 941 t
ORIGIN

```

```

Query Match 84.0%; Score 16.8; DB 6: Length 5451;
Best Local Similarity 90.0%; Pred. No. 4.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ctaccggcgccgaaggccggg 20
    |||||||
Db 2049 CTACCGCGCGGAGGCCCTGG 2068

```

RESULT 13

```

LOCUS AB019552 6220 bp DNA linear BCT 18-JAN-2002
DEFINITION Aeropyrum pernix genes for 16S rRNA, ITS, 23S rRNA, partial and
complete sequence, strain:OH2.
ACCESSION AB019552
VERSION AB019552.2 GI:18249697
KEYWORDS Aeropyrum pernix (strain:OH2) DNA.
SOURCE Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
Desulfurococcaceae; Aeropyrum.
ORGANISM

```

```

REFERENCE 1 (sites)
AUTHORS Nomura, N., Sako, Y., Morinaga, Y., Kogishi, T. and Uchida, A.
TITLE Intraspetic genetic polymorphism in the rRNA gene locus of the
hyperthermophilic archaeon Aeropyrum pernix, implying the presence
of multiple hotspots for intron homing
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 6220)
AUTHORS Nomura, N.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1998) Norimichi Nomura, Kyoto University, Lab. of

```


Marine Microbiology, Graduate School of Agriculture,
Kitashirakawa-Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan
(E-mail:j54718@sakura.kudpc.kyoto-u.ac.jp, Tel:81-75-753-6219,
Fax:81-75-753-6226)

COMMENT On Jan 18, 2002 this sequence version replaced gi:3869179.

FEATURES
source
1..6220
/organism="Aeropyrum pernix"
/strain="OH2"
/db_xref="taxon:56636"
298..1798
/product="16S ribosomal RNA"
1799..2155
/note="internal transcribed spacer (ITS) region"
2156..5238
/product="23S ribosomal RNA"
BASE COUNT 1088 a 1878 c 2370 g 884 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 1; Length 6220;
Best Local Similarity 90.0%; Pred. No. 4e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccgcgcgaagccggg 20
||| || ||||| |||||
Db 3812 CTAACCTGCCGAGCGGG 3831

RESULT 14
AB078015 6220 bp DNA linear BCT 19-JAN-2002
LOCUS
DEFINITION Aeropyrum pernix genes for 16S rRNA, 23S rRNA and ITS region,
complete sequence, strain:OH1.
ACCESSION AB078015
VERSION
KEYWORDS AB078015.1 GI:18250956
SOURCE
ORGANISM Aeropyrum pernix (strain:OH1) DNA.
Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
Desulfurococcaceae; Aeropyrum.

REFERENCE 1 (sites)
AUTHORS Nomura,N., Morinaga,Y., Kogishi,T., Kim,E., Sako,Y. and Uchida,A.
TITLE Heterogeneous yet partially similar introns reside in identical
positions of the rRNA genes in natural isolates of the archaeon
Aeropyrum pernix
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 6220)
AUTHORS Nomura,N.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-2002) Norimichi Nomura, Kyoto University,
Division of Applied Biosciences, Graduate School of Agriculture,
Sakyo, Kyoto 606-8502, Japan
(E-mail:j54718@sakura.kudpc.kyoto-u.ac.jp, Tel:81-75-753-6224,
Fax:81-75-753-6226)

FEATURES
source
1..6220
/organism="Aeropyrum pernix"
/strain="OH1"
/db_xref="taxon:56636"
298..1797
/gene="arnS"
/product="16S ribosomal RNA"
298..1797
/gene="arnS"
misc_feature 1798..2155
/note="internal transcribed spacer (ITS) region"
2156..5238
rRNA
/gene="arnL"
/product="23S ribosomal RNA"
2156..5238
gene
/gene="arnL"
BASE COUNT 1086 a 1877 c 2372 g 885 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 1; Length 6220;
Best Local Similarity 90.0%; Pred. No. 4e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ctaccgcgcgaagccggg 20
||| || ||||| |||||
Db 3812 CTAACCTGCCGAGCGGG 3831

RESULT 15
AB078016 6220 bp DNA linear BCT 19-JAN-2002
LOCUS
DEFINITION Aeropyrum pernix genes for 16S rRNA, 23S rRNA and ITS region,
complete sequence, strain:OH3.
ACCESSION AB078016
VERSION
KEYWORDS AB078016.1 GI:18250957
SOURCE
ORGANISM Aeropyrum pernix (strain:OH3) DNA.
Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
Desulfurococcaceae; Aeropyrum.

REFERENCE 1 (sites)
AUTHORS Nomura,N., Morinaga,Y., Kogishi,T., Kim,E., Sako,Y. and Uchida,A.
TITLE Heterogeneous yet partially similar introns reside in identical
positions of the rRNA genes in natural isolates of the archaeon
Aeropyrum pernix
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 6220)
AUTHORS Nomura,N.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-2002) Norimichi Nomura, Kyoto University,
Division of Applied Biosciences, Graduate School of Agriculture,
Sakyo, Kyoto 606-8502, Japan
(E-mail:j54718@sakura.kudpc.kyoto-u.ac.jp, Tel:81-75-753-6224,
Fax:81-75-753-6226)

FEATURES
source
1..6220
/organism="Aeropyrum pernix"
/strain="OH3"
/db_xref="taxon:56636"
298..1798
/gene="arnS"
/product="16S ribosomal RNA"
298..1798
/gene="arnS"
misc_feature 1799..2155
/note="internal transcribed spacer (ITS) region"
2156..5238
rRNA
/gene="arnL"
/product="23S ribosomal RNA"
2156..5238
gene
/gene="arnL"
BASE COUNT 1091 a 1877 c 2368 g 884 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 1; Length 6220;
Best Local Similarity 90.0%; Pred. No. 4e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ctaccgcgcgaagccggg 20
||| || ||||| |||||
Db 3812 CTAACCTGCCGAGCGGG 3831

Search completed: June 19, 2002, 15:43:19
Job time: 13457 sec

Result No.	Score	Query Match	Length	DB	ID	Description
1	20	100.0	44	20	AA02689	CFDV-DNA fragment
2	20	100.0	44	20	AA02687	CFDV-DNA fragment
3	16.8	84.0	5451	22	AAF26319	Pseudomonas sp lip
4	16	80.0	341	22	AAH68205	Human cervical can
5	15.8	79.0	2040	22	AAH68825	Mycobacterium tube
C	15.8	79.0	2388	22	AAD16357	Human sbq34976IGBA
C	15.8	79.0	2238	22	AAI166145	PAM polynucleotide
C	15.8	79.0	3217	22	AAI166138	PAM encoding cDNA
C	15.8	79.0	3226	12	AAQ11038	Encodes Rat C-term

Region	note=
misc binding	38 43

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted and is derived by analysis of the total score distribution.

```
FT      /*tag= f
FT      /note= "Region binds to nucleotides 27 to 32"
PN
XX
XX
PD      DE19730502-A1.
XX
XX      21-JAN-1999.
XX
XX      16-JUL-1997; 97DE-1030502.
XX
XX      16-JUL-1997; 97DE-1030502.
XX
XX      (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX      Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX
XX      WPI; 1999-096863/09.
XX
XX      Coconut foliar decay virus promoters - for gene expression in
FT      bacteria and yeasts
XX
XX      Disclosure; Fig 4; 14pp; German.
XX
XX      This invention describes a coconut foliar decay virus (CFDV) DNA fragment
CC      that includes the stem-loop structure of CFDV DNA but lacks the
CC      translation start codons of open reading frames ORF1 and/or ORF2. The
CC      new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
CC      promoter for tissue-specific (especially phloem-specific) gene expression
CC      in plants and for production of chimeric constructs for transient or
CC      stable expression. Certain fragments of CFDV DNA have stronger promoter
CC      activity in E. coli than the CamV 35S promoter.
XX
XX      Sequence 44 BP; 8 A; 16 C; 17 G; 3 T; 0 other;
SQ
Query Match      100.0%; Score 20; DB 20; Length 44;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ctaccggccgaagccggg 20
DB      |||||||||||||||
        24 ctaccggccgaagccggg 43

RESULT      2
AAAX02687
ID      AAAX02687 standard; DNA; 44 BP.
XX
XX      AAX02687;
XX
XX      10-MAY-1999 (first entry)
XX
XX      CFDV-DNA fragment stem loop repeat region.
XX
XX      Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
KW      yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.
XX
XX      Coconut foliar decay virus
XX
XX      Key      Location/Qualifiers
FH      stem_loop      4..22
FT      /*tag= a
FT      /note= "RPT1"
FT      4..10
FT      /*tag= b
FT      /note= "Region binds to nucleotides 16 to 22"
FT      16..22
FT      /*tag= c
FT      /note= "Region binds to nucleotides 4 to 10"
FT      27..43
FT      /*tag= d
FT      /note= "RPT2"
FT      27..32
FT      /*tag= e

FT      /note= "Region binds to nucleotides 38 to 43"
FT      /*tag= f
FT      /note= "Region binds to nucleotides 27 to 32"
PN
XX
XX      DE19730535-A1.
XX
XX      21-JAN-1999.
XX
XX      16-JUL-1997; 97DE-1030535.
XX
XX      16-JUL-1997; 97DE-1030535.
XX
XX      (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX      Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX
XX      WPI; 1999-096867/09.
XX
XX      Coconut foliar decay virus promoters - for gene expression in
FT      plants, bacteria and yeasts
XX
XX      Disclosure; Fig 4; 12pp; German.
XX
XX      This invention describes a coconut foliar decay virus (CFDV) DNA fragment
CC      that includes the stem-loop structure of CFDV DNA but lacks the
CC      translation start codons of open reading frames ORF1 and/or ORF2. The
CC      new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
CC      promoter for tissue-specific (especially phloem-specific) gene expression
CC      in plants and for production of chimeric constructs for transient or
CC      stable expression. Certain fragments of CFDV DNA have stronger promoter
CC      activity in E. coli than the camV 35S promoter.
XX
XX      Sequence 44 BP; 8 A; 16 C; 17 G; 3 T; 0 other;
SQ
Query Match      100.0%; Score 20; DB 20; Length 44;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ctaccggccgaagccggg 20
DB      |||||||||||||||
        24 ctaccggccgaagccggg 43

RESULT      3
AAAF26319
ID      AAF26319 standard; DNA; 5451 BP.
XX
XX      AAF26319;
XX
XX      02-MAY-2001 (first entry)
XX
XX      Pseudomonas sp lipopeptide synthase DNA ORF08563.
XX
XX      Lipopeptide synthase; transgenic plant; antiviral; antibacterial;
KW      antifungal; surfactant; microbial depollution; emulsifier;
KW      tertiary crude oil recovery; ORF08563; ds.
XX
XX      Pseudomonas sp.
XX
XX      WO200107589-A2.
XX
XX      01-FEB-2001.
XX
XX      21-JUL-2000; 2000WO-EP07002.
XX
XX      27-JUL-1999; 99DE-1035106.
XX
XX      (TIGR-) TIGR INST GENOMIC RES.
XX      (QIAG-) QIAGEN GMBH.
XX      (GBF-) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX      (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
```

(MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.

Fraser CM, Tuemmler B, Hoheisel J, Duesterhoeft A, Hilbert H;
Timmis KN, Moore E, Straetz M, Heim S, Golyshin P;
WPI: 2001-159715/16.

New DNA encoding a bacterial lipopeptide synthase, useful e.g. for
producing antimicrobial agents, surfactants with increased
bioavailability in microbial depollution methods, and transgenic plants

Claim 3a; Page 28-30; 33pp; German.

This invention describes a novel DNA sequence (I) that expresses products
having the biological function of lipopeptide synthase. The invention
also describes (1) recombinant expression vectors containing (I); (2)
prokaryotic and eukaryotic cells transformed or transfected with (I) or
the vector of (2); (3) production of lipopeptide synthases by culturing
cells of (2); (4) expression products (II) of (I), and synthetic proteins
or peptides with the same sequences; (5) mono- or poly-clonal antibodies
(Ab) specific for (II); (6) hybridoma cells that produce monoclonal Ab;
and (7) transgenic plants that contain cells of (2), (I), and their
fragments, are useful for expression of recombinant lipopeptide
synthases, and as probes and primers for detection, isolation and
amplification of full-length cDNA sequences. (I) are used to produce
transgenic plants. Lipopeptide synthases are useful for production of
antiviral, antibacterial or antifungal lipopeptides, particularly useful
for treatment of plants, and for production of lipopeptide surfactants
used for increasing bioavailability in microbial depollution processes
(acting as emulsifiers) and in tertiary crude oil recovery. Individual
subunits from different lipopeptide synthases may be combined to allow
synthesis of many different biologically active substances.

Sequence 5451 BP; 937 A; 1766 C; 1807 G; 941 T; 0 other;

Query Match 84.0%; Score 16.8; DB 22; Length 5451;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccgcgcgaagccggg 20
|||||||
Db 2049 ctaccgcgcgaagccgtgg 2068

RESULT 4
AAH68825
ID AAH68825 standard; cDNA; 341 BP.

XX AC AAH68825;
XX DT 19-SEP-2001 (first entry)
XX DE Human cervical cancer marker nucleic acid 99.

XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200142467-A2.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US33312.

XX PR 08-DEC-1999; 99US-0169681.

XX PR 21-DEC-1999; 99US-0171350.

XX PR 14-MAR-2000; 2000US-0189315.

XX PR 12-MAY-2000; 2000US-0203791.

XX PR 09-JUN-2000; 2000US-0210600.

XX PR 21-JUL-2000; 2000US-0220114.

XX

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Deeds J, Berger A, Zhao X;

WPI: 2001-375006/39.

New isolated nucleic acid for diagnosing and treating cervical cancer
and for assessing and detecting compounds for treating the cancer -

Claim 1; Page 120; 1051pp; English.

The invention relates to novel genes (AAH68727-AAH73383) associated with
cervical cancer with cytostatic activity. The nucleic acids and encoded
polypeptides are useful: to assess if a patient is afflicted with
cervical cancer or has a pre-malignant condition; to monitor the
progression of cervical cancer or a premalignant condition in a patient;
and to select and/or assess the efficacy of a compound or therapy for
inhibiting cervical cancer in a patient. The nucleic acids may also be
useful for gene therapy.

Sequence 341 BP; 72 A; 93 C; 93 G; 68 T; 15 other;

Query Match 80.0%; Score 16; DB 22; Length 341;

Best Local Similarity 94.1%; Pred. No. 3.2e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ccgcgcgcgaagccggg 20

|||||||

Db 192 ccgcgcgcgaagccggg 208

RESULT 5

AAH52030

ID AAH52030 standard; DNA; 2040 BP.

XX AC AAH52030;

XX DT 04-SEP-2001 (first entry)

XX DE Mycobacterium tuberculosis potential drug target gene SEQ ID 84.

XX KW Drug target; growth; organism viability; characterisation; ds.

XX OS Mycobacterium tuberculosis.

XX PN WO200135317-A1.

XX PD 17-MAY-2001.

XX PF 13-NOV-2000; 2000WO-US31152.

XX PR 12-NOV-1999; 99US-0165086.

XX PR 12-NOV-1999; 99US-0165124.

XX PR 01-FEB-2000; 2000US-0179531.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Eisenberg D, Rotstein SH, Marcotte EM;

XX WPI: 2001-329193/34.

XX P-PSDB; AAG81179.

XX PT Identifying nucleotide or polypeptide sequence for use as drug target,
involves providing algorithm that analyzes a functional relationship
between nucleotide or polypeptide sequences, and comparing the
sequences -

XX PS Disclosure; Page 103-104; 207pp; English.

XX CC This invention relates to a method for identifying a nucleotide or
polypeptide sequence that may be a drug target, or essential for growth
or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092

CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analysing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism.
 XX
 SQ Sequence 2040 BP; 374 A; 673 C; 671 G; 322 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 2040;
 Best Local Similarity 89.5%; Pred. No. 3.5e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taccggcggaagccggg 20
 ||||| ||||| |||||
 Db 1731 taccggcggaagccggg 1749

RESULT 6
 AAD16357/C
 ID AAD16357 standard; DNA; 2238 BP.

XX AC AAD16357;

XX DT 19-NOV-2001 (first entry)

XX DE Human sbg34976IGBa gene #1.

XX Human; Alzheimer's disease; anyotropic lateral sclerosis;
 KW ALS; Zollinger-Ellison syndrome; immune system disease; schizophrenia;
 KW inflammation; haematopoietic disease; anxiety; feeding disorder; aging;
 KW anorexia; depression; cardiovascular disease; sleep disorder; seizure;
 KW memory alteration; migraine; stroke; asthma; neuropathy; hypoglycaemia;
 KW sexual disorder; growth abnormality; infection; autoimmune disease;
 KW rheumatoid arthritis; cataractogenesis; angiogenesis; atherosclerosis;
 KW cerebral ischaemia; cirrhosis; Huntington's disease; Hodgson's disease;
 KW hypercholesterolaemia; headache; amnesia; cardiac arrhythmia; obesity;
 KW diabetes mellitus; glomerulonephritis; renovascular hypertension;
 KW cancer; vaccine; gene therapy; sbg34976IGBa gene; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX CDS 1..2238
 XX FT /*tag= a
 XX FT /product= "Human sbg34976IGBa protein #1"

XX WO200160850-A1.

XX PD 23-AUG-2001.

XX PF 14-FEB-2001; 2001WO-US04703.

XX PR 14-FEB-2000; 2000US-0182172.

XX PR 29-FEB-2000; 2000US-0185084.

XX PR 18-APR-2000; 2000US-0198583.

XX PR 04-OCT-2000; 2000US-0237963.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM PLC.

XX PI Agarwal P, Kabnick KS, Murdock PR, Rizvi SK, Smith RF, Xiang Z;

XX WPI; 2001-536566/59.

XX P-PSDB; AA09450.

XX New secreted and membrane associated polypeptides for treating
 PT Alzheimer's disease, psoriasis, cancer, enterocolitis, sleep and sexual

PT disorders, stroke, and asthma -
 XX Claim 2; Page 48-49; 94pp; English.
 XX The present sequence is a gene encoding human sbg34976IGBa protein,
 CC a membrane bound protein of the invention.
 CC The invention relates to secreted and membrane associated polypeptides
 CC and nucleic acid molecules encoding such polypeptides. Sequences of the
 CC invention are useful for treating diseases such as Alzheimer's disease,
 CC amyotrophic lateral sclerosis (ALS), Zollinger-Ellison syndrome, diseases
 CC of the immune system, haematopoietic disease, inflammation, anxiety,
 CC schizophrenia, feeding disorders, anorexia, depression, social, sexual
 CC and rewarded behaviour, cardiovascular disease, sleep disorder, learning
 CC and memory alteration and altered immune response, seizure, migraine,
 CC cancer, stroke, asthma, neuropathy, aging, sexual disorders, treatment
 CC of transsexuals, growth abnormalities, obesity, infections, autoimmune
 CC diseases (e.g. rheumatoid arthritis), cataractogenesis, angiogenesis,
 CC disorders associated with healthy maintenance of gastric mucosa and
 CC repair of acute and chronic mucosal lesion, lung carcinoma, cerebral
 CC ischaemia, atherosclerosis, cirrhosis, Huntington's disease, headache,
 CC amnesia, multiple sclerosis, Hodgson's disease, hypercholesterolaemia,
 CC congestive heart failure, cardiac arrhythmias, hypercholesterolaemia,
 CC viral and non-viral hepatitis, type I and type II diabetes mellitus,
 CC glomerulonephritis, renovascular hypertension, hypoglycaemia, periodic
 CC paralyseis, tendinitis and malignant hyperthermia. Polypeptides of the
 CC invention are used to identify membrane bound and soluble receptors.
 CC They are also useful as vaccines for inducing an immunological response
 CC in a mammal. Polynucleotides of the invention are used in gene therapy.
 CC They are also valuable for chromosome localisation studies and tissue
 CC expression studies.
 XX
 SQ Sequence 2238 BP; 400 A; 744 C; 743 G; 351 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 2238;
 Best Local Similarity 89.5%; Pred. No. 3.4e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taccggcggaagccggg 20
 ||||| ||||| |||||
 Db 1684 TACCGCGCGCAGCGCGC 1666

RESULT 7
 AAI66145/C
 ID AAI66145 standard; cDNA; 2388 BP.

XX AC AAI66145;

XX DT 14-JAN-2002 (first entry)

XX DE PAM polynucleotide 3.

XX Chinese hamster ovary; PAM; ECI.14.17.3; ss;
 KW Peptidylglycine alpha-amidating monooxygenase.

XX OS Unidentified.

XX CNL167154-A.

XX PD 10-DEC-1997.

XX PF 16-APR-1997; 97CN-0106382.

XX PR 16-APR-1997; 97CN-0106382.

XX (SHAN-) SHANGHAI BIO-CHEM INST CHINESE ACAD SCI.

XX Li B, Jiang Z, Xia Q;

XX WPI; 2001-590349/67.

XX Amidase gene and modifying application of its expressed active products

PT to polypeptide amidating -

XX Example 1; Fig 2; 37pp; Chinese.

XX The invention relates to a rat amidase gene, peptidylglycine
CC alpha-amidating monooxygenase/PAM (EC1.14.17.3) and its active expression
CC product in Chinese hamster ovary (CHO) cell. The invention uses in situ
CC hybridization and a polymerase chain reaction (PCR) method to screen and
CC obtain three rPAM gene fragments, then through technologies of point
CC mutation and PCR recombination to splice rPAM holoenzyme gene and form
CC its eucaryon expression plasmid so as to obtain the active expression in
CC CHO cell. The obtained expression product can be directly used for
CC amidation modification of polypeptide and protein. The present sequence
CC is that of a polynucleotide of the PAM gene (AAI66138), useful to the
XX invention.

XX Sequence 2388 BP; 638 A; 549 C; 589 G; 612 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 2388;
Best Local Similarity 89.5%; Pred. No. 3.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggcggaagggcg 19
||||||| ||||| ||

Db 1823 CTACCGGGCGAAGGCTGG 1805

RESULT 8
AAI66138/C
ID AAI66138 standard; cDNA; 3217 BP.

XX AAI66138;

XX 14-JAN-2002 (first entry)

XX PAM encoding cDNA.

XX Chinese hamster ovary; PAM;

XX Peptidylglycine alpha-amidating monooxygenase; ss.

XX Unidentified.

XX Key Location/Qualifiers
CDS 11..2629
FT /*tag= a
FT /product= "PAM"
FT /EC_number= "1.14.17.3"

XX CNL167154-A.

XX 10-DEC-1997.

XX 16-APR-1997; 97CN-0106382.

XX 16-APR-1997; 97CN-0106382.

XX (SHAN-) SHANGHAI BIO-CHEM INST CHINESE ACAD SCI.

XX Li B, Jiang Z, Xia Q;

XX WPI: 2001-590349/67.

XX P-PSDB; AAM51684.

XX Amidase gene and modifying application of its expressed active products
XX to polypeptide amidating -

XX Claim 2; Fig 1; 37pp; Chinese.

XX The invention relates to a rat amidase gene, peptidylglycine
CC alpha-amidating monooxygenase/PAM (EC1.14.17.3) and its active expression
CC product in Chinese hamster ovary (CHO) cell. The invention uses in situ
CC hybridization and a polymerase chain reaction (PCR) method to screen and

CC obtain three rPAM gene fragments, then through technologies of point
CC mutation and PCR recombination to splice rPAM holoenzyme gene and form
CC its eucaryon expression plasmid so as to obtain the active expression in
CC CHO cell. The obtained expression product can be directly used for
CC amidation modification of polypeptide and protein.

XX Sequence 3217 BP; 839 A; 762 C; 791 G; 825 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 3217;

Best Local Similarity 89.5%; Pred. No. 3.4e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggcggaagggcg 19
||||||| ||||| ||

Db 2652 CTACCGGGCGAAGGCTGG 2634

RESULT 9
AAQ11038/C
ID AAQ11038 standard; cDNA; 3226 BP.

XX AAQ11038;

XX 17-MAY-1991 (first entry)

XX Encodes Rat C-terminal amidating enzyme.

XX C-terminal amidating enzyme; calcitonin; gastrin; ss.

XX Rattus.

XX Key Location/Qualifiers

FT CDS 2..2563

FT /*tag= a

FT /product= C-terminal amidating enzyme

XX WO9102790-A.

XX 07-MAR-1991.

XX 14-AUG-1990; 90WO-JP01036.

XX 15-AUG-1989; 89JP-0209687.

XX 31-OCT-1989; 89JP-0281933.

XX 26-MAR-1990; 90JP-0076331.

XX 24-APR-1990; 90JP-0106412.

XX 02-AUG-1990; 90JP-0205475.

XX (SHIS) SHISEIDO KK.

XX WPI: 1991-087274/12.

XX P-PSDB; AAR11110.

XX New enzymes useful in peptide C-amidation - convert peptide C-
XX terminal glycine to alpha hydroxy glycine then remove this leaving
XX peptide C-term amide.

XX Disclosure: Fig 6; 114pp; Japanese.

XX The enzyme encoded by this sequence is involved in the C-terminal
XX amidation reactions. It can be used to amide physiologically
XX active peptides such as calcitonin and gastrin. It also amidates D-
XX Tyr-Trp-Gly, Phe-Gly-Phe-Gly and Gly-Phe-Gly. The region from
XX position 2 to 630 was used as a probe to isolate the C-terminal
XX amidating enzyme from the horse.
XX See also AAQ11039-Q11042.

XX Sequence 3226 BP; 845 A; 748 C; 788 G; 844 T; 1 other;

Query Match 79.0%; Score 15.8; DB 12; Length 3226;

Best Local Similarity 89.5%; Pred. No. 3.3e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccgccggaagccgg 19
 ||||| ||||| ||

Db 2585 CTACCCGGCGAGGCTGG 2567

RESULT 10

AAQ05637/C
 ID AAQ05637 standard; DNA; 3586 BP.

XX AC AAQ05637;

XX DT 14-DEC-1990 (first entry)

XX DE Sequence which hybridises with DNA encoding alpha amidating enzyme.

XX KW Alpha amidating enzyme; peptidyl amide; ss.

XX FH Key Location/Qualifiers
 CDS 346..2907

XX FT /*tag= a

XX PN EP382403-A.

XX PD 16-AUG-1990.

XX PF 01-FEB-1990; 90EP-0301034.

XX PR 06-FEB-1989; 89US-0307366.

XX PA (UNIG-) UNIGENE LAB INC.

XX PI Betelsen AH, Mehta NM, Beaudry GA;

XX DR WPI; 1990-248308/33.

XX DR P-PSDB; AAR96426.

XX PT Prodn. of alpha amidating enzyme - using prokaryotic or
 eukaryotic expression vectors having transcriptional promoter

XX PS Claim 2; page 17; 24pp; English.

XX CC This sequence hybridises, under stringent conditions, with a DNA
 sequence (I) which encodes an alpha amidating enzyme (AEE) and is
 foreign to the host cell into which it is transformed. (I)
 is connected to a promoter and is contained in an expression
 vector. The AEE encoded is used as a catalyst in the conversion
 of a peptidyl substrate to a corresp. peptidyl amide. This can be
 used for making a protein biologically active, e.g. calcitonin or
 growth hormone releasing factor. See also AAQ05638.

XX SQ Sequence 3586 BP; 908 A; 879 C; 906 G; 893 T; 0 other;

Query Match 79.0%; Score 15.8; DB 11; Length 3586;

Best Local Similarity 89.5%; Pred. NO. 3.3e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccgccggaagccgg 19

||||| ||||| ||

Db 2929 CTACCCGGCGAGGCTGG 2911

RESULT 11

AAQ10277/C
 ID AAQ10277 standard; DNA; 3586 BP.

XX AC AAQ10277;

XX DT 08-APR-1991 (first entry)

XX DE Type A alpha-amidating enzyme.

XX

KW Amidation; post-translational modification; AE; ss.

XX OS Rattus rattus.

XX FH Key Location/Qualifiers

CDS 346..2907

FT /*tag= a

FT /label= AE Type A

XX PN AU9049043-A.

XX PD 29-NOV-1990.

XX PF 01-FEB-1990; 90AU-0049043.

XX PR 06-FEB-1989; 89US-0307336.

XX PA (UNIG-) UNIGENE LAB INC.

XX PI Betelsen AH, Mehta NM, Beaudry GA;

XX DR WPI; 1991-022433/04.

XX DR P-PSDB; AAR10322.

XX PT Expression systems for amidating enzyme - comprises prokaryotic
 or eukaryotic hosts contg. a recombinant expression vector contg.

XX PT the corresp. DNA sequence.

XX PS Disclosure; Page 2; 52pp; English.

XX CC The DNA was isolated from libraries prepd. from RNA extracted from
 rat medullary thyroid carcinoma tissues or derived cell lines such
 as IVI 10028. The cDNAs extracted could be divided into distinct
 CC types, the sequence below being of Type A. (Type B is given in
 CC AAQ10278). The sequence can be inserted into expression vectors for
 CC the prodn. of recombinant alpha-amidating enzyme (AE) useful for
 CC post translational modification of other recombinant polypeptides
 CC such as calcitonin. The sequence includes a membrane spanning
 CC domain which may be undesirable in a recombinant DNA expression
 CC system, possibly inactivating the enzyme. A stop codon can be
 CC placed upstream of the domain-encoding sequence, pref. between
 CC posns. 2025 and 2275 (counting from the CDS). This results in the
 CC expression of a 75 rather than 94 kD. The sequence of Type A and
 CC Type B are substantially identical with the exception of an intron
 CC region from bases 1178-1492 (CDS) of the Type B cDNA.

XX SQ Sequence 3586 BP; 909 A; 879 C; 906 G; 892 T; 0 other;

Query Match 79.0%; Score 15.8; DB 12; Length 3586;

Best Local Similarity 89.3%; Pred. NO. 3.3e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccgccggaagccgg 19

||||| ||||| ||

Db 2929 CTACCCGGCGAGGCTGG 2911

RESULT 12

AAQ21277/C

ID AAS21277 standard; cDNA; 4185 BP.

XX AC AAS21277;

XX DT 24-OCT-2001 (first entry)

XX DE Human cDNA sequence encoding for PRO4329 polypeptide.

XX KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.

XX OS Homo sapiens.
 XX PN WO2001040466-A2.
 XX PD 07-JUN-2001.
 XX PF 01-DEC-2000; 2000WO-US32678.
 XX PR 01-DEC-1999; 99WO-US28301.
 XX PR 01-DEC-1999; 99WO-US28634.
 XX PR 02-DEC-1999; 99WO-US28551.
 XX PR 02-DEC-1999; 99WO-US28564.
 XX PR 02-DEC-1999; 99WO-US28565.
 XX PR 09-DEC-1999; 99US-0170262.
 XX PR 16-DEC-1999; 99WO-US30095.
 XX PR 20-DEC-1999; 99WO-US30911.
 XX PR 20-DEC-1999; 99WO-US30999.
 XX PR 30-DEC-1999; 99WO-US31243.
 XX PR 06-JAN-2000; 2000WO-US00277.
 XX PR 06-JAN-2000; 2000WO-US00376.
 XX PR 11-FEB-2000; 2000WO-US03565.
 XX PR 18-FEB-2000; 2000WO-US04341.
 XX PR 22-FEB-2000; 2000WO-US04342.
 XX PR 24-FEB-2000; 2000WO-US04914.
 XX PR 24-FEB-2000; 2000WO-US05004.
 XX PR 01-MAR-2000; 2000WO-US05601.
 XX PR 20-MAR-2000; 2000WO-US07377.
 XX PR 21-MAR-2000; 2000WO-US07532.
 XX PR 30-MAR-2000; 2000WO-US08439.
 XX PR 17-MAY-2000; 2000WO-US13705.
 XX PR 22-MAY-2000; 2000WO-US14042.
 XX PR 30-MAY-2000; 2000WO-US14941.
 XX PR 02-JUN-2000; 2000WO-US15264.
 XX PR 10-NOV-2000; 2000WO-US30873.
 XX PA (GETH) GENENTECH INC.
 XX PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
 XX PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 XX PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX WPI: 2001-408281/43.
 XX P-PSDB; AAU12205.
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical -
 XX Claim 3; Fig 67; 813pp; English.
 XX AA521244-AA521518 encode for novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.

XX SQ Sequence 4185 BP; 713 A; 1363 C; 1307 G; 802 T; 0 other;
 XX Query Match 79.0%; Score 15.8; DB 22; Length 4185;
 XX Best Local Similarity 89.5%; Pred. No. 3.3e+02;
 XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 taccgcgcgcgaagcgccgg 20
 Db 1820 TACCCGCGCGCAGCGCGG 1802
 RESULT 13
 AAS59517/c
 ID AAS59517 standard; DNA: 18796 BP.
 XX AC AAS59517;
 XX DT 13-FEB-2002 (first entry)
 XX DE Propionibacterium acnes immunogenic protein encoding DNA #12.
 XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 XX KW dermatological; osteopathic; neuroprotectant; ds.
 XX OS Propionibacterium acnes.
 XX PN WO200181581-A2.
 XX PD 01-NOV-2001.
 XX PF 20-APR-2001; 2001WO-US12865.
 XX PR 21-APR-2000; 2000US-199047P.
 XX PR 02-JUN-2000; 2000US-208841P.
 XX PR 07-JUL-2000; 2000US-216747P.
 XX PA (CORI-) CORIXA CORP.
 XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX WPI: 2001-616774/71.
 PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX Claim 1; SEQ ID No 12; 1069pp; English.
 XX Sequences AAS59506-AAS59804 represent DNA molecules encoding
 CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
 CC associated DNA sequences are used in the treatment, prevention and
 CC diagnosis of medical conditions caused by P. acnes. The disorders include
 CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
 CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
 CC in infections of bone, joints and the central nervous system, however it
 CC is particularly involved in the inflammatory lesions associated with acne
 CC vulgaris. A method for detecting the presence or absence of P. acnes in a
 CC patient comprises contacting a sample with a binding agent that binds to
 CC the proteins of the invention and determining the amount of bound protein
 CC in the sample. The polypeptides may be used as antigens in the production
 CC of antibodies specific for P. acnes proteins. These antibodies can be
 CC used to downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
 CC polypeptides shown in AAU2344-42548 and AAU67479.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 18796 BP; 3824 A; 5690 C; 5570 G; 3711 T; 1 other;

SQ Query Match 77.0%; Score 15.4; DB 23; Length 18796;
Best Local Similarity 94.1%; Pred. No. 4.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cccgcgcgaagccggg 20

||||| |||||||

DB 9479 CCCGACGAGGCCGGG 9463

RESULT 14

AA136348/C

ID AA136348 standard; DNA; 19820 BP.

XX

AC AA136348;

XX

DT 08-JAN-2002 (first entry)

XX

DE Human musculoskeletal system related polynucleotide SEQ ID NO 2713.

XX

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
antibacterial; antifungal; antiparasitic;
cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
neurological disease; infection; human; secreted protein;
musculoskeletal system; ds.

OS Homo sapiens.

XX

PN WO200155367-A1.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01338.

XX

31-JAN-2000; 2000US-0179065.

PR

04-FEB-2000; 2000US-0180628.

PR

24-FEB-2000; 2000US-0184664.

PR

02-MAR-2000; 2000US-0186350.

PR

16-MAR-2000; 2000US-0189874.

PR

17-MAR-2000; 2000US-0190076.

PR

18-APR-2000; 2000US-0198123.

PR

19-MAY-2000; 2000US-0205515.

PR

07-JUN-2000; 2000US-0209467.

PR

28-JUN-2000; 2000US-0214886.

PR

30-JUN-2000; 2000US-0215135.

PR

07-JUL-2000; 2000US-0216647.

PR

07-JUL-2000; 2000US-0216880.

PR

11-JUL-2000; 2000US-0217487.

PR

14-JUL-2000; 2000US-0217496.

PR

26-JUL-2000; 2000US-0220963.

PR

26-JUL-2000; 2000US-0220964.

PR

14-AUG-2000; 2000US-0224518.

PR

14-AUG-2000; 2000US-0224519.

PR

14-AUG-2000; 2000US-0225213.

PR

14-AUG-2000; 2000US-0225214.

PR

14-AUG-2000; 2000US-0225266.

PR

14-AUG-2000; 2000US-0225267.

PR

14-AUG-2000; 2000US-0225268.

PR

14-AUG-2000; 2000US-0225270.

PR

14-AUG-2000; 2000US-0225447.

PR

14-AUG-2000; 2000US-0225757.

PR

14-AUG-2000; 2000US-0225758.

PR

14-AUG-2000; 2000US-0225759.

PR

18-AUG-2000; 2000US-0226279.

PR

22-AUG-2000; 2000US-0226681.

PR

22-AUG-2000; 2000US-0226686.

PR

22-AUG-2000; 2000US-0227182.

PR

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241321.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-451937/48.

XX Isolated polypeptide for treating, preventing and/or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -

XX Example 2; SEQ ID NO 2713; 781pp + Sequence Listing; English.

XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (AB03087-AB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The genes are isolated from a range of human
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 19820 BP; 6182 A; 3722 C; 4043 G; 5871 T; 2 other;

XX Query Match 77.0%; Score 15.4; DB 22; Length 19820;
 XX Best Local Similarity 94.1%; Pred. No. 4.4e+02;
 XX Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cccgcgcgaagccggcgg 20

Db 817 CCCGCCGAGGCGGG 801

RESULT 15
 AAF30757/c
 ID AAF30757 standard; DNA; 47981 BP.
 XX
 AC AAF30757;
 XX
 DT 21-JUN-2001 (first entry)
 XX
 DE Micromonospora megalomicea megalomicin biosynthetic gene cluster.
 XX Micromonospora megalomicea subsp. nigra.
 KW Megalomicin; meg gene; polyketide synthase; antibiotic;
 KW motilide; antiparasitic; ds.
 XX
 OS Micromonospora megalomicea subsp. nigra.
 PH Key Location/Qualifiers
 FT CDS complement (1..144)
 FT /*tag= a
 FT /*partial
 FT /gene= "megBVI(megT)"
 FT /product= "TDP-4-keto-6-deoxyglucose-
 FT 2,3-dehydratase"
 FT /note= "encodes AAB82201"
 FT 928..2061
 FT /*tag= b
 FT /gene= "megDVI"
 FT /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
 FT /note= "encodes AAB82202"
 FT 2072..3382
 FT /*tag= c
 FT /gene= "megDI"
 FT /product= "TDP-megosamine glycosyltransferase"
 FT /note= "eryCIII homologue; encodes AAB802203"
 FT 3462..4634
 FT /*tag= d
 FT /gene= "megY"
 FT /product= "mycarose O-acyltransferase"
 FT /note= "encodes AAB82204"
 FT 4651..5775
 FT /*tag= e
 FT /gene= "megDII"
 FT /product= "TDP-3-keto-6-deoxyhexose
 FT 3-aminotransaminase"
 FT /note= "eryCI, DnrJ homologue, encodes AAB82205"
 FT 5822..6595
 FT /*tag= f
 FT /gene= "megDIII"
 FT /product= "daunosaminyl-N,N-dimethyltransferase"
 FT /note= "eryCVI homologue; encodes AAB82206"
 FT 6592..7197
 FT /*tag= g
 FT /gene= "megDIV"
 FT /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
 FT /note= "eryVIII, dnmU homologue, encodes AAB82207"
 FT 7220..8206
 FT /*tag= h
 FT /gene= "megDV"
 FT /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
 FT /note= "eryVIV, dnmV homologue; encodes AAB82208"
 FT complement (8228..9220)
 FT /*tag= i
 FT /gene= "megBII-1(megDVII)"
 FT /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
 FT /note= "encodes AAB82209"
 FT complement (9226..10479)
 FT /*tag= j
 FT /gene= "megBV"
 FT /product= "TDP-mycarose glycosyltransferase"
 FT /note= "encodes AAB82210"
 FT complement (10483..11424)
 FT /*tag= k
 FT

```

FT FT /gene= "megBIV"
FT FT /product= "NDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT FT /note= "encodes AAB82211"
FT FT /gene= "megA"
FT FT /tag= l
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT /note= "polyketide synthase 1"
FT FT /tag= m
FT FT /gene= "megA"
FT FT /function= "AT-L"
FT FT /tag= o
FT FT /gene= "megA"
FT FT /function= "KSL"
FT FT /tag= p
FT FT /gene= "megA"
FT FT /function= "AT1"
FT FT /tag= q
FT FT /gene= "megA"
FT FT /function= "KR1"
FT FT /tag= r
FT FT /gene= "megA"
FT FT /function= "ACPI"
FT FT /tag= s
FT FT /gene= "megA"
FT FT /function= "KS2"
FT FT /tag= t
FT FT /gene= "megA"
FT FT /function= "AT2"
FT FT /tag= u
FT FT /gene= "megA"
FT FT /function= "KR2"
FT FT /tag= v
FT FT /gene= "megA"
FT FT /function= "ACP2"
FT FT /tag= w
FT FT /gene= "megAII"
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT /note= "polyketide synthase 2"
FT FT /tag= x
FT FT /gene= "megAII"
FT FT /function= "KS3"
FT FT /tag= y
FT FT /gene= "megAII"
FT FT /function= "AT3"
FT FT /tag= z
FT FT /gene= "megAII"
FT FT /function= "KR3 (inactive)"
FT FT /tag= aa
FT FT /gene= "megAII"
FT FT /function= "ACP3"
FT FT /tag= ab

```

```

FT FT /gene= "megAII"
FT FT /function= "KS4"
FT FT /tag= ac
FT FT /gene= "megAII"
FT FT /function= "AT4"
FT FT /tag= ad
FT FT /gene= "megAII"
FT FT /function= "DH4"
FT FT /tag= ae
FT FT /gene= "megAII"
FT FT /function= "ER4"
FT FT /tag= af
FT FT /gene= "megAII"
FT FT /function= "KR4"
FT FT /tag= ag
FT FT /gene= "megAII"
FT FT /function= "ACP4"
FT FT /tag= ah
FT FT /gene= "megAII"
FT FT /product= "megalomicin 6-deoxyerythronolide B
FT FT /note= "polyketide synthase 3"
FT FT /tag= ai
FT FT /gene= "megAII"
FT FT /function= "KS5"
FT FT /tag= aj
FT FT /gene= "megAII"
FT FT /function= "AT5"
FT FT /tag= ak
FT FT /gene= "megAII"
FT FT /function= "KR5"
FT FT /tag= al
FT FT /gene= "megAII"
FT FT /function= "ACP5"
FT FT /tag= am
FT FT /gene= "megAII"
FT FT /function= "KS6"
FT FT /tag= an
FT FT /gene= "megAII"
FT FT /function= "AT6"
FT FT /tag= ao
FT FT /gene= "megAII"
FT FT /function= "KR6"
FT FT /tag= ap

```

Query Match 77.0%; Score 15.4; DB 22; Length 47981;
 Best Local Similarity 94.1%; Pred. No. 4.1e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cccggccgaagccggg 20
 |||||
 Db 6886 CCCGCCGAAGGTCGGG 6870

Search completed: June 19, 2002, 16:50:59
 Job time: 17517 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:47:08 ; search time 247.21 seconds
(without alignments)
19.872 Million cell updates/sec

Title: US-09-462-955B-1_COPY_682_701

Perfect score: 20

Sequence: 1 ctaccggcggaagccggg 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2.6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2.6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2.6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2.6/ptodata/1/ina/6B_COMB.seq:*
- 5: /cgn2.6/ptodata/1/ina/PCTUS_COMB.seq:*
- 6: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	21	4	US-09-462-975-7
2	20	100.0	1291	4	US-09-462-975-1
3	15.8	79.0	3226	2	US-08-070-301-10
4	15.4	77.0	4411529	4	US-09-103-840A-1
5	15.2	76.0	1520	1	US-08-225-477B-1
6	15.2	76.0	1520	5	PCT-US95-04353-1
7	15.2	76.0	50937	4	US-09-428-517-1
8	14.8	74.0	254	3	US-08-866-340-31
9	14.8	74.0	254	4	US-09-103-875-37
10	14.8	74.0	1002	1	US-08-457-707A-9
11	14.8	74.0	1002	1	US-08-812-025-9
12	14.8	74.0	1002	1	US-09-138-873A-9
13	14.8	74.0	3490	2	US-08-841-483-3
14	14.8	74.0	3490	4	US-09-382-911-3
15	14.8	74.0	4094	2	US-08-841-483-5
16	14.8	74.0	4094	4	US-09-382-911-5
17	14.8	74.0	4403765	4	US-09-103-840A-2
18	14.4	72.0	485	2	US-08-403-852D-15
19	14.4	72.0	485	3	US-08-510-646B-15
20	14.4	72.0	485	4	US-09-231-818-15
21	14.4	72.0	752	1	US-08-725-182C-1
22	14.4	72.0	752	3	US-09-013-150-1
23	14.4	72.0	752	4	US-09-359-503-1
24	14.4	72.0	752	4	US-09-062-422-1
25	14.4	72.0	752	4	US-09-396-184-1
26	14.4	72.0	752	4	US-08-937-263B-1
27	14.4	72.0	755	1	US-08-791-495-8

c	28	14.4	72.0	1879	6	5212296-5	Patent No. 5212296
	29	14.2	71.0	33	2	US-08-365-486A-2	Sequence 2, Appli
	30	14.2	71.0	33	4	US-08-880-342-2	Sequence 2, Appli
c	31	14.2	71.0	696	4	US-09-220-528-1	Sequence 1, Appli
	32	14.2	71.0	696	4	US-09-220-528-2	Sequence 2, Appli
c	33	14.2	71.0	1652	4	US-09-220-528-68	Sequence 68, Appli
	34	14.2	71.0	1652	4	US-09-220-528-69	Sequence 69, Appli
	35	14	70.0	3744	2	US-08-348-353-16	Sequence 16, Appli
	36	14	70.0	3744	2	US-08-465-965-16	Sequence 16, Appli
	37	14	70.0	3744	3	US-08-465-966-16	Sequence 16, Appli
c	38	13.8	69.0	23	1	US-08-308-949A-12	Sequence 12, Appli
	39	13.8	69.0	23	4	US-09-462-975-6	Sequence 6, Appli
c	40	13.8	69.0	212	1	US-08-435-684A-32	Sequence 32, Appli
	41	13.8	69.0	212	2	US-08-934-877A-32	Sequence 32, Appli
c	42	13.8	69.0	212	4	US-08-871-678C-32	Sequence 32, Appli
	43	13.8	69.0	423	1	US-08-470-179-171	Sequence 171, App
c	44	13.8	69.0	595	1	US-08-580-038-25	Sequence 25, Appli
	45	13.8	69.0	774	2	US-08-471-371-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-462-975-7
; Sequence 7, Application US/09462975
; Patent No. 6303345
; GENERAL INFORMATION:
; APPLICANT: Rohde, Wolfgang
; APPLICANT: Becker, Dieter
; APPLICANT: Randles, John W.
; APPLICANT: Hehn, Alain
; APPLICANT: Salamini, Francesco
; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER
; FILE REFERENCE: 23232.000301
; CURRENT APPLICATION NUMBER: US/09/462.975
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/EP98/04345
; PRIOR FILING DATE: 1998-07-13
; PRIOR APPLICATION NUMBER: 19730502.4
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: /note-synthetic
; OTHER INFORMATION: construct
US-09-462-975-7

Query Match 100.0%; Score 20; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctaccggcggaagccggg 20
Db 1 ctaccggcggaagccggg 20

RESULT 2
US-09-462-975-1
; Sequence 1, Application US/09462975
; Patent No. 6303345
; GENERAL INFORMATION:
; APPLICANT: Rohde, Wolfgang
; APPLICANT: Becker, Dieter
; APPLICANT: Randles, John W.
; APPLICANT: Hehn, Alain
; APPLICANT: Salamini, Francesco
; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER

FILE REFERENCE: 23232.0003U1
CURRENT APPLICATION NUMBER: US/09/462,975
FILING DATE: 2000-05-17
PRIOR APPLICATION NUMBER: PCT/EP98/04345
PRIOR FILING DATE: 1998-07-13
PRIOR APPLICATION NUMBER: 19730502.4
PRIOR FILING DATE: 1997-07-16
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1291
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:/note=synthetic
OTHER INFORMATION: construct
US-09-462-975-1

Query Match 100.0%; Score 20; DB 4; Length 1291;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccgcccgaagccggg 20
|||||
Db 1072 ctaccgcccgaagccggg 1091

RESULT 3
US-08-070-301-10/c
Sequence 10, Application US/08070301
Patent No. 5871995
GENERAL INFORMATION:
APPLICANT: IIDA, Toshio
APPLICANT: KAMINUMA, Toshihiko
APPLICANT: FUSE, Yuka
APPLICANT: TAJIMA, Masahiro
APPLICANT: YANAGI, Mitsuo
APPLICANT: OKAMOTO, Hiroshi
APPLICANT: KISHIMOTO, Jiro
APPLICANT: IFUKU, Ohji
APPLICANT: KATO, Ichiro
TITLE OF INVENTION: ENZYME PARTICIPATING IN C-TERMINAL
AMIDATION, AND METHOD OF PREPARING SAME AND USE THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wegner, Cantor, Mueller & Player, P.C.
STREET: 1233 20th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-8218
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/070,301
FILING DATE: 24-MAY-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1-209687
FILING DATE: 15-AUG-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1-181933
FILING DATE: 31-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-76331
FILING DATE: 26-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-106412
FILING DATE: 24-APR-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-205475
FILING DATE: 02-AUG-1990
ATTORNEY/AGENT INFORMATION:
NAME: Player, William E.
REGISTRATION NUMBER: 31,409
REFERENCE/DOCKET NUMBER: P-450-22830
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-040
TELEFAX: (202) 835-0605
TELEX: 440706
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 3226 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
ORIGINAL SOURCE:
ORGANISM: Rat
FEATURE:
NAME/KEY: CDS
LOCATION: 2..831
US-08-070-301-10

Query Match 79.0%; Score 15.8; DB 2; Length 3226;
Best Local Similarity 89.5%; Pred. No. 64;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccgcccgaagccggg 19
|||||
Db 2585 ctaccgcccgaagccggg 2567

RESULT 4
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 77.0%; Score 15.4; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cccgcccgaagccggg 19
|||||
Db 1558394 cccgcccgaagccggg 1558409

RESULT 5
US-08-225-477B-1
Sequence 1, Application US/08225477B
Patent No. 5635370
GENERAL INFORMATION:


```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-103-875-37

Query Match      74.0%; Score 14.8; DB 4; Length 254;
Best Local Similarity 88.9%; Pred No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 accgcgcccgaagccggg 20
   ||| ||| ||| ||| ||| |||
Db 66 ACCAGGCCCAAGGCCGGG 49

RESULT 10
US-08-457-797A-9
; Sequence 9, Application US/08457797A
; Patent No. 5689045
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Transgenic pathogen-resistant organism
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue and Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,797A
; FILING DATE: June 1, 1995
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 18,839
; REFERENCE/DOCKET NUMBER: A29542 FWC - 37/31335
; TELEPHONE: (212) 408-2500
; TELEFAX: (212) 765-2519
; TELEX: 238555
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Hordeum vulgare
; STRAIN: L.
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..63
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 64..861
; OTHER INFORMATION: /codon_start= 64
; OTHER INFORMATION: /function= "chitinase"
; OTHER INFORMATION: /product= "26 kD preprotein of chitinase G (ChIG)"
; OTHER INFORMATION: /note= "antifungal activity, especially on
; OTHER INFORMATION: Trichoderma reesei and Fusarium sporotrichoides as
; OTHER INFORMATION: well as Rhizoctonia solani and Botrytis cinerea."
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 862..1002

```

```

; US-08-866-340-31/c
; Sequence 31, Application US/08866340
; Patent No. 6020318
; GENERAL INFORMATION:
; APPLICANT: Szyf, Moshe
; APPLICANT: Bigey, Pascal
; APPLICANT: Ramchandani, Shyam
; TITLE OF INVENTION: DNA METHYLTRANSFERASE GENOMIC
; TITLE OF INVENTION: SEQUENCES AND ANTISENSE OLIGONUCLEOTIDES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States of America
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,340
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Keown, Wayne A.
; REGISTRATION NUMBER: 33,923
; REFERENCE/DOCKET NUMBER: 106.101.187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-866-340-31

Query Match      74.0%; Score 14.8; DB 3; Length 254;
Best Local Similarity 88.9%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 accgcgcccgaagccggg 20
   ||| ||| ||| ||| ||| |||
Db 66 ACCAGGCCCAAGGCCGGG 49

RESULT 9
US-09-103-875-37/c
; Sequence 37, Application US/09103875A
; Patent No. 6221849
; GENERAL INFORMATION:
; APPLICANT: Szyf, Moshe
; APPLICANT: Bigey, Pascal
; APPLICANT: Ramchandani, Shyam
; TITLE OF INVENTION: DNA METHYLTRANSFERASE GENOMIC SEQUENCES AND ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES
; FILE REFERENCE: 106101.194
; CURRENT APPLICATION NUMBER: US/09/103,875A
; CURRENT FILING DATE: 1998-06-24
; EARLIER APPLICATION NUMBER: 60/069,865
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: 08/866,340
; EARLIER FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 138

```



```

; OTHER INFORMATION: /partial
; OTHER INFORMATION: /note= "11 nucleotides at 3' end not shown"
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 905..910
; OTHER INFORMATION: /note= "potential polyadenylation
; OTHER INFORMATION: signal"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 64..294
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 298..312
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 349..378
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 466..588
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 607..861
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 133..861
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; US-08-457-797A-9

```

```

Query Match 74.0%; Score 14.8; DB 1; Length 1002;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 3 accgcgcgaagccggg 20
    ||| ||||| |||||
Db 677 ACCGGCGCGAGCGCGG 694

```

```

RESULT 11
US-08-812-025-9
; Sequence 9, Application US/08812025
; Patent No. 5804184
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Transgenic pathogen-resistant organism
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue and Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/812,025
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Tenser, Arthur
; REGISTRATION NUMBER: 18,839

```

```

; NAME: Kole, Lisa
; REGISTRATION NUMBER: 35,225
; REFERENCE/DOCKET NUMBER: A29542 FWC - 37/31335
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 408-2500
; TELEFAX: (212) 765-2519
; TELEX: 238555
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Hordeum vulgare
; STRAIN: L.
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..63
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 64..861
; OTHER INFORMATION: /codon_start= 64
; OTHER INFORMATION: /function= "chitinase"
; OTHER INFORMATION: /product= "26 kD preprotein of chitinase G (Ch1G)"
; OTHER INFORMATION: /note= "antifungal activity, especially on
; OTHER INFORMATION: Trichoderma reesei and Fusarium sporotrichoides as
; OTHER INFORMATION: well as Rhizoctonia solani and Botrytis cinerea."
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 862..1002
; OTHER INFORMATION: /partial
; OTHER INFORMATION: /note= "11 nucleotides at 3' end not shown"
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 905..910
; OTHER INFORMATION: /note= "potential polyadenylation
; OTHER INFORMATION: signal"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 64..294
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 298..312
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 349..378
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 466..588
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 607..861
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 133..861
; OTHER INFORMATION: /note= "probable signal peptide
; OTHER INFORMATION: sequence"
; US-08-812-025-9

```

```

Query Match 74.0%; Score 14.8; DB 1; Length 1002;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Thu Jun 20 06:56:49 2002

LOCATION: 64..294
OTHER INFORMATION: /note= "probable signal peptide"
OTHER INFORMATION: sequence"

FEATURE:
NAME/KEY: sig_peptide
LOCATION: 298..312
OTHER INFORMATION: /note= "probable signal peptide"
OTHER INFORMATION: sequence"

FEATURE:
NAME/KEY: sig_peptide
LOCATION: 349..378
OTHER INFORMATION: /note= "probable signal peptide"
OTHER INFORMATION: sequence"

FEATURE:
NAME/KEY: sig_peptide
LOCATION: 466..588
OTHER INFORMATION: /note= "probable signal peptide"
OTHER INFORMATION: sequence"

FEATURE:
NAME/KEY: sig_peptide
LOCATION: 607..861
OTHER INFORMATION: /note= "probable signal peptide"
OTHER INFORMATION: sequence"

FEATURE:
NAME/KEY: mat_peptide
LOCATION: 133..861
US-09-138-873A-9

Query Match 74.0%; Score 14.8; DB 4; Length 1002;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 acccgccgaagccggg 20
DB 677 ACCGGCCGCGAGCCGG 694

RESULT 13

US-08-841-483-3/c
; Sequence 3, Application US/08841483B
; Patent No. 5976875
; GENERAL INFORMATION:
; APPLICANT: Prescott, Steven M.
; APPLICANT: Bunting, Michaeline
; APPLICANT: Tang, Wen
; APPLICANT: Topham, Matthew
; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
; TITLE OF INVENTION: Methods of Use Thereof
; FILE REFERENCE: 2037.2.1a
; CURRENT APPLICATION NUMBER: US/08/841,483B
; CURRENT FILING DATE: 1997-04-22
; EARLIER APPLICATION NUMBER: 60/016,210
; EARLIER FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3490
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (89)..(2875)
US-08-841-483-3

Query Match 74.0%; Score 14.8; DB 2; Length 3490;
Best Local Similarity 88.9%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 acccgccgaagccggg 20
DB 3255 ACCGGCCGCGAGCCAGG 3238

QY 3 acccgccgaagccggg 20
DB 677 ACCGGCCGCGAGCCGG 694

RESULT 12

US-09-138-873A-9

; Sequence 9, Application US/09138873A

; Patent No. 6271438

; GENERAL INFORMATION:

; APPLICANT: Transgenic pathogen-resistant organism

; TITLE OF INVENTION: Transgenic pathogen-resistant organism

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Baker & Botts

; STREET: 30 Rockefeller Plaza

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10112

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/138,873A

; FILING DATE: August 24, 1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Tenser, Arthur

; REGISTRATION NUMBER: 18,839

; NAME: Kote, Lisa

; REGISTRATION NUMBER: 35,225

; REFERENCE/DOCKET NUMBER: A29542 FWC - 37/31335

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 408-2500

; TELEFAX: (212) 765-2519

; TELEX: 238555

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1002 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; ORIGINAL SOURCE:

; ORGANISM: Hordeum vulgare

; STRAIN: L.

; FEATURE:

; NAME/KEY: 5'UTR

; LOCATION: 1..63

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 64..861

; OTHER INFORMATION: /codon_start= 64

; OTHER INFORMATION: /function= "chitinase"

; OTHER INFORMATION: /product= "26 kD preprotein of chitinase G (ChicG)"

; OTHER INFORMATION: /note= "antifungal activity, especially on

; OTHER INFORMATION: Trichoderma reesei and Fusarium sporotrichoides as

; OTHER INFORMATION: well as Rhizoctonia solani and Botrytis cinerea."

; FEATURE:

; NAME/KEY: 3'UTR

; LOCATION: 862..1002

; OTHER INFORMATION: /partial

; OTHER INFORMATION: /note= "11 nucleotides at 3' end not shown"

; FEATURE:

; NAME/KEY: polyA_signal

; LOCATION: 905..910

; OTHER INFORMATION: /note= "potential polyadenylation

; OTHER INFORMATION: signal"

; FEATURE:

; NAME/KEY: sig_peptide

RESULT 14
 US-09-382-911-3/c
 : Sequence 3, Application US/09382911
 : Patent No. 6221658
 : GENERAL INFORMATION:
 : APPLICANT: Prescott, Steven M.
 : APPLICANT: Bunting, Michaeline
 : APPLICANT: Tang, Wen
 : APPLICANT: Topham, Matthew
 : TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
 : TITLE OF INVENTION: Methods of Use Thereof
 : FILE REFERENCE: 2037.2.1a
 : CURRENT APPLICATION NUMBER: US/09/382,911
 : CURRENT FILING DATE: 1999-08-25
 : PRIOR APPLICATION NUMBER: 08/841,483
 : PRIOR FILING DATE: 1997-04-22
 : PRIOR APPLICATION NUMBER: 60/016,210
 : PRIOR FILING DATE: 1996-04-22
 : NUMBER OF SEQ ID NOS: 33
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 3
 : LENGTH: 3490
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: (89)..(2875)
 US-09-382-911-3

Query Match 74.0%; Score 14.8; DB 4; Length 3490;
 Best Local Similarity 88.9%; Pred. No. 1.7e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 acccgccggaagccggg 20
 ||||| ||||| |||||
 Db 3255 ACCCGCCGAAGACCAGG 3238

RESULT 15
 US-08-841-483-5/c
 : Sequence 5, Application US/08841483B
 : Patent No. 5976875
 : GENERAL INFORMATION:
 : APPLICANT: Prescott, Steven M.
 : APPLICANT: Bunting, Michaeline
 : APPLICANT: Tang, Wen
 : APPLICANT: Topham, Matthew
 : TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
 : TITLE OF INVENTION: Methods of Use Thereof
 : FILE REFERENCE: 2037.2.1a
 : CURRENT APPLICATION NUMBER: US/08/841,483B
 : CURRENT FILING DATE: 1997-04-22
 : EARLIER APPLICATION NUMBER: 60/016,210
 : EARLIER FILING DATE: 1996-04-22
 : NUMBER OF SEQ ID NOS: 33
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 5
 : LENGTH: 4094
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: (126)..(3479)
 US-08-841-483-5

Query Match 74.0%; Score 14.8; DB 2; Length 4094;
 Best Local Similarity 88.9%; Pred. No. 1.7e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 acccgccggaagccggg 20
 ||||| ||||| |||||
 Db 3859 ACCCGCCGAAGACCAGG 3842

Search completed: June 19, 2002, 15:49:00
 Job time: 13798 sec

us-09-462-955b-1_copy_682_701.rni

Thu Jun 20 06:56:49 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 14:04:08 ; Search time 7489.97 Seconds
(without alignments)
36.040 Million cell updates/sec

Title: US-09-462-955b-1_COPY_682_701

Perfect score: 20

Sequence: 1 ctaccgcggcgaagccggg 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estab:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	17	85.0	448	9	BE114197
2	17	85.0	491	9	BE114196
3	16.8	84.0	612	10	BE238940
4	16.4	82.0	786	10	BF679019
5	16.4	82.0	849	9	AL539606
6	16.4	82.0	996	10	BE260627
7	16.4	82.0	1172	10	BM467932
8	16	80.0	490	10	BG320293
9	15.8	79.0	309	10	BF286795
10	15.8	79.0	313	12	B18042
11	15.8	79.0	318	10	BF779841
12	15.8	79.0	319	10	BF779842
13	15.8	79.0	422	9	AW142153
14	15.8	79.0	426	10	BG298646
15	15.8	79.0	438	10	BM376491
16	15.8	79.0	470	12	B18122
17	15.8	79.0	473	10	BG417064

c	18	15.8	79.0	507	9	AW919607
	19	15.8	79.0	552	10	BG846215
	20	15.8	79.0	575	10	BF103096
c	21	15.8	79.0	669	12	AG093531
	22	15.8	79.0	678	10	BG846216
	23	15.8	79.0	717	10	BE454137
	24	15.8	79.0	876	10	BG181627
c	25	15.8	79.0	877	10	B1908420
c	26	15.8	79.0	909	12	AZ210881
	27	15.8	79.0	920	12	CNS02261
	28	15.8	79.0	980	10	BF780277
	29	15.8	79.0	1006	12	CNS01TTB
	30	15.8	79.0	1156	10	BE791915
	31	15.8	79.0	1214	12	AG086086
	32	15.4	77.0	147	9	AA283327
	33	15.4	77.0	232	9	BB571165
c	34	15.4	77.0	352	9	AW337668
	35	15.4	77.0	377	9	BB813476
	36	15.4	77.0	416	9	AA997260
c	37	15.4	77.0	417	10	BM199463
	38	15.4	77.0	438	9	AW619475
	39	15.4	77.0	478	9	AW916542
c	40	15.4	77.0	483	9	BB697893
	41	15.4	77.0	500	12	BH353262
c	42	15.4	77.0	617	12	BH549318
c	43	15.4	77.0	640	10	BM439271
c	44	15.4	77.0	685	10	B1559944
45	15.4	77.0	717	9	A1963704	

ALIGNMENTS

RESULT 1

BE114197
LOCUS
DEFINITION
UI-R-CAO-axo-b-10-0-UI.s1 UI-R-CAO Rattus norvegicus cDNA clone
UI-R-CAO-axo-b-10-0-UI 3', mRNA sequence.
BE114197
VERSION
KEYWORDS
SOURCE
ORGANISM
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS
TITLE
Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL
MEDLINE
COMMENT
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dt track not found. Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: Clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 forward
FOLYA=NO.
Location/Qualifiers
1. .448
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CAO-axo-b-10-0-UI"
/clone_lib="UI-R-CAO"
/lab_host="DH10B (Life Technologies)"

FEATURES
source

/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CAO library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

Genome Research 6:791-806, 1996)
TAG_SEQ=None found"

BASE COUNT 69 a 125 c 171 g 81 t 2 others
ORIGIN

Query Match 85.0%; Score 17; DB 9; Length 448;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cccgcccgaagccggg 20
|||||

Db 32 CCCGCCGAGGCCGGG 48

RESULT 2
LOCUS BE114196 491 bp mRNA linear EST 13-JUN-2000
DEFINITION UI-R-CAO-axo-b-09-0-UI.s1 UI-R-CAO Rattus norvegicus cDNA clone
UI-R-CAO-axo-b-09-0-UI 3', mRNA sequence.

ACCESSION BE114196
VERSION BE114196
KEYWORDS BE114196.1 GI:8506301
SOURCE EST.

ORGANISM Norway rat.
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 491)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477

COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: mscares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward

POLYA-No.

FEATURES Location/Qualifiers

1..491
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CAO-axo-b-09-0-UI"
/clone_lib="UI-R-CAO"

/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CAO library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares,

Genome Research 6:791-806, 1996)

TAG_SEQ=None found"
BASE COUNT 77 a 142 c 179 g 92 t 1 others
ORIGIN

Query Match 85.0%; Score 17; DB 9; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cccgcccgaagccggg 20
|||||

Db 32 CCCGCCGAGGCCGGG 48

RESULT 3

LOCUS BE238940/c

DEFINITION MD0596 Meloidogyne incognita J2 (#MD99-1) Meloidogyne incognita
cDNA clone 3438 5' similar to SEC-2 protein (Y09293), mRNA
sequence.

ACCESSION BE238940

VERSION BE238940.1 GI:9033904

KEYWORDS EST.

SOURCE southern root-knot nematode.

ORGANISM Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

REFERENCE 1 (bases 1 to 612)

AUTHORS Dautova,M., Gommers,F.J., Bakker,J. and Smant,G.

TITLE 5' end expressed sequence tags from Meloidogyne incognita
preparasitic J2 cDNA library

JOURNAL Unpublished (2000)

COMMENT Contact: Smant G / Dautova M

Laboratory of Nematology

Wageningen University and Research Center

Box 8123, Binnenhaven 10, 6709 PD Wageningen, The Netherlands

Tel: 31 317 485 254

Fax: 31 317 484 254

Email: Geert.Smant@medew.nema.wau.nl,
Makedonka.Dautova@medew.nema.wau.nl

Insert Length: 612 Std Error: 0.00

Seq primer: T7 promoter primer

High quality sequence stop: 612.

FEATURES

source

1..612
/organism="Meloidogyne incognita"

/db_xref="taxon:6306"

/clone="3438"

/clone_lib="Meloidogyne incognita J2 (#MD99-1)"
/dev_stage="second stage parasitic juveniles (J2)"

/note="Vector: pMAK1; Site_1: Sfi IA; Site_2: Sfi IB; cDNA was synthesized using SMART III oligo, CDS III oligo d(T)
30 (Clontech) and Superscript II reverse transcriptase
(Life Technology). cDNA clones were size fractionated and
directionally ligated in the Sfi IA restriction site at
5' end and Sfi IB at 3' end of pMAK1. pMAK1 was derived
from the plasmid pCDNA II (Invitrogen)."

BASE COUNT 219 a 146 c 120 g 127 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 612;

Best Local Similarity 90.0%; Pred. No. 1.6e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggccgaagccggg 20

|||||

Db 470 CTACGGCGCGAGGCCGGG 451

RESULT 4

BF679019

LOCUS
DEFINITION 602153663F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4294839 5',
EST. 786 bp mRNA linear EST 21-DEC-2000

ACCESSION BF679019
VERSION BF679019.1 GI:11952914
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 786)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI144 row: m column: 16
High quality sequence stop: 564.

FEATURES
Location/Qualifiers
1..786
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4294839"
/lab_host="NIH_MGC_83"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site 1: Sfil (ggcgctcgcc); Site 2: Sfil (ggcattatggcc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATATGCCC-3' and 3' adaptor
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
218 a 174 c 167 g 227 t

BASE COUNT
ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 786;
Best Local Similarity 94.4%; Pred. No. 2.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 acccgccgaagccggg 20
|||||
DB 703 ACCCGCGGAGGCGGG 720

RESULT 5
AL539606
LOCUS
DEFINITION AL539606 LTI_FL013_FBRn1 Homo sapiens cDNA clone CSODF036YD18 5
prime, mRNA sequence.
ACCESSION AL539606
VERSION AL539606.1 GI:12868977
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 849)
Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Location/Qualifiers
1..849
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODF036YD18"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DHL08"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 189 a 187 c 267 g 205 t 1 others
ORIGIN

Query Match 82.0%; Score 16.4; DB 9; Length 849;
Best Local Similarity 85.0%; Pred. No. 2.4e+03;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccgcccgaagccggg 20
|||||
DB 412 CGACGCGCGGCGGCGG 431

RESULT 6
BE260627/c
LOCUS
DEFINITION BE260627 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3161405 5',
mRNA sequence.
ACCESSION BE260627
VERSION BE260627.1 GI:9132066
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 996)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI19 row: c column: 06
High quality sequence stop: 4
High quality sequence start: 661.

FEATURES
Location/Qualifiers
1..996
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3161405"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DHL08 (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 246 a 271 c 263 g 216 t
ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 996;
Best Local Similarity 94.4%; Pred. No. 2.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 taccgcgcgaagccgg 19
||||| ||||| ||||| |||||

Db 760 TACCCGCCAAGCCGG 743

RESULT 7
BM467932 1172 bp mRNA linear EST 05-FEB-2002
LOCUS
DEFINITION AGENCOURT_6437854 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532832
5', mRNA sequence.

ACCESSION BM467932
VERSION
KEYWORDS
SOURCE EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1172)
AUTHORS NTH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov

Tissue Procurement: AFCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12216 row: 1 column: 17
High quality sequence start: 5
High quality sequence stop: 638.

FEATURES
Location/Qualifiers
1..1172
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5532832"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1; NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb."
268 a 384 c 299 g 221 t

BASE COUNT 268 a 384 c 299 g 221 t
ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 1172;
Best Local Similarity 94.4%; Pred. No. 2.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 accgcgcgaagccgg 20
||||| ||||| ||||| |||||

Db 940 ACTCGCCGAAGCCGG 957

RESULT 8
BG320293/c 490 bp mRNA linear EST 27-FEB-2001
LOCUS
DEFINITION Zm03_12e11_A Zm03_AAFc_ECORC_cold_stressed_maize_seedlings Zea mays

cdNA clone Zm03_12e11, mRNA sequence.
ACCESSION BG320293
VERSION BG320293.1 GI:13149971
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 490)
Simmonds, J.A., Singh, J.A., Piche, C., Cass, L., Couroux, P., De Moors, A., Harris, L.J., Hattori, J.I., Ouellet, T., Robert, L.S., Sprott, D., and Tinker, N.A.

Expressed Sequence Tags from Cold-Stressed Maize Seedlings Grown Under High Light Intensity
Unpublished (2001)
Contact: Singh, J.A.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
960 Carling Avenue, Bldg. 20, Ottawa, Ontario, K1A 0C6, Canada
Tel: (613) 759-1662
Fax: (613) 759-1701
Email: singhia@em.agr.ca.

FEATURES
Location/Qualifiers
1..490
/organism="Zea mays"
/cultivar="C0328"
/db_xref="taxon:4577"
/clone="Zm03_12e11"
/clone_lib="Zm03_AAFc_ECORC_cold_stressed_maize_seedlings"
/tissue_type="Leaf, crown"
/dev_stage="4-leaf"
/note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: Eco RI; Site 2: Xho I; Corn seedlings at 4-leaf stage were exposed to low temperature/high light (100C/700-800uE/m2/s) for 4 days. Plants were grown/treated by J. Simmonds/L. Cass. Library prepared by C. Piche using Stratagene kit."
56 a 168 c 171 g 71 t 24 others

BASE COUNT 56 a 168 c 171 g 71 t 24 others
ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 490;
Best Local Similarity 80.0%; Pred. No. 3.4e+03;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccgcgcgaagccgg 20
||||| ||||| ||||| |||||

Db 135 CCACCCGCCGATGCCGG 116

RESULT 9
BF286795/c 309 bp mRNA linear EST 28-NOV-2000
LOCUS
DEFINITION EST451386 Rat Gene Index, normalized rat, Rattus norvegicus CDNA
Rattus norvegicus cDNA clone RGIF578, mRNA sequence.

ACCESSION BF286795
VERSION BF286795.1 GI:11217865
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 309)
AUTHORS Malek, R.L., Cho, J., Lee, Y., Karamycheva, S., Parviz, B., Perlea, G., Sultana, R., Tsai, J., White, J., Quackenbush, J., and Lee, N.H.
TITLE Generation of ESTs from Normalized Rat Embryo, Bonto Soares Unpublished (2000)
JOURNAL Other ESTs: EST451385
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.

FEATURES

source

Location/Qualifiers
1. 309
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RCIFS78"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"
/note="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROV, RBR, RKI, RLI, RPL, REM, RMU, RSP
' RHE, RCP, RPN"
58 a 92 c 69 g 90 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 79.0%; Score 15.8; DB 10; Length 309;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggccgaagccgg 19

Db 47 CTACCCGGCGAAGGCTGT 29

RESULT 10

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

B18042 A-237G3-TV CIT978SK Homo sapiens genomic clone A-237G3, DNA
sequence.
B18042 B18042.1 GI:2121509
GSS.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 313)
Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.
Use of a BAC End Sequence Database for Sequence-Ready Map Building
Unpublished (1997)
Other_GSSs: 237G3.TP
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: 17
Class: BAC ends.
Location/Qualifiers
1. 313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="A-237G3"
/clone_lib="CIT978SK"
/sex="Female"
/cell_type="Fibroblast"
/note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
Caltech Human BAC Library A"
41 a 119 c 112 g 40 t 1 others

FEATURES

source

Query Match
Best Local Similarity 79.0%; Score 15.8; DB 12; Length 313;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

BASE COUNT

ORIGIN

Query Match
Best Local Similarity 79.0%; Score 15.8; DB 10; Length 318;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 12

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

B1779842 B1779842.1 GI:15782734
GSS.
human.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 318)
Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardie,L.,
Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
Development of Barley Transcriptome Resources
Unpublished (2001)
Contact: Waugh R
Unit of Genomics
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: rwaugh@scri.sari.ac.uk
All sequence has a phred quality score of 20 or over
Seq primer: M13 reverse.
Location/Qualifiers
1. 318
/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="EBem07_S0001_C21"
/clone_lib="IGF Barley EBem07 library"
/tissue_type="Embryo"
/dev_stage="28 days post anthesis"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from embryos dissected from developing grains (28
days post anthesis) in glasshouse grown barley plants.
Developed as part of the barley transcriptome resources of
BBSRC/SEERAD funded cereal IGF (Investigating Gene
Function) project."
59 a 92 c 87 g 80 t

Best Local Similarity 89.5%; Pred. No. 4.1e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggccgaagccgg 19

Db 182 CTGCCCCGGCGAAGGCTGT 164

RESULT 11

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

B1779841 B1779841.1 GI:15782733
GSS.
barley.
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 318)
Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardie,L.,
Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
Development of Barley Transcriptome Resources
Unpublished (2001)
Contact: Waugh R
Unit of Genomics
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: rwaugh@scri.sari.ac.uk
All sequence has a phred quality score of 20 or over
Seq primer: M13 reverse.
Location/Qualifiers
1. 318
/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="EBem07_S0001_C21"
/clone_lib="IGF Barley EBem07 library"
/tissue_type="Embryo"
/dev_stage="28 days post anthesis"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from embryos dissected from developing grains (28
days post anthesis) in glasshouse grown barley plants.
Developed as part of the barley transcriptome resources of
BBSRC/SEERAD funded cereal IGF (Investigating Gene
Function) project."
59 a 92 c 87 g 80 t

FEATURES

source

Query Match
Best Local Similarity 79.0%; Score 15.8; DB 10; Length 318;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggccgaagccgg 19

Db 33 CTACCCGGCGAAGGCTGT 51

RESULT 12

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

B1779842 B1779842.1 GI:15782734
GSS.
barley.
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 318)
Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardie,L.,
Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
Development of Barley Transcriptome Resources
Unpublished (2001)
Contact: Waugh R
Unit of Genomics
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: rwaugh@scri.sari.ac.uk
All sequence has a phred quality score of 20 or over
Seq primer: M13 reverse.
Location/Qualifiers
1. 318
/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="EBem07_S0001_C21"
/clone_lib="IGF Barley EBem07 library"
/tissue_type="Embryo"
/dev_stage="28 days post anthesis"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from embryos dissected from developing grains (28
days post anthesis) in glasshouse grown barley plants.
Developed as part of the barley transcriptome resources of
BBSRC/SEERAD funded cereal IGF (Investigating Gene
Function) project."
59 a 92 c 87 g 80 t

Thu Jun 20 06:56:49 2002

```

SOURCE
ORGANISM
    barley.
    Hordeum vulgare
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
    ; Triticeae; Hordeum.
REFERENCE
AUTHORS
    Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
    Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
TITLE
    Development of Barley Transcriptome Resources
JOURNAL
    Unpublished (2001)
COMMENT
    Contact: Waugh R
    Unit of Genomics
    Scottish Crop Research Institute
    Invergowrie, Dundee, DD2 5DA, Scotland, UK
    Tel: 00 44 1382 562731
    Fax: 00 44 1382 562426
    Email: rwaugh@scri.sari.ac.uk
    All sequence has a Phred quality score of 20 or over
    Seq primer: M13 reverse
FEATURES
    source
        1..319
            Location/Qualifiers
                /organism="Hordeum vulgare"
                /cultivar="Optic"
                /db_xref="taxon:4513"
                /clone="EBem07_S0001_C22"
                /clone_lib="IGF Barley EBem07 library"
                /tissue_type="Embryo"
                /dev_stage="28 days post anthesis"
                /lab_host="DH10B"
                /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                Non-normalised library, directionally cloned into pSPORT1.
                Derived from embryos dissected from developing grains (28
                days post anthesis) in glasshouse grown barley plants.
                Developed as part of the barley transcriptome resources of
                BBSRC/SEERAD funded cereal IGF (Investigating Gene
                Function) project."
BASE COUNT
    61 a 91 c 86 g 81 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 10; Length 319;
Best Local Similarity 89.5%; Pred. No. 4.1e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggcgagcgccg 19
    ||||| ||||| |||||
Db 34 CTACCGCGCGAATGCCGG 52

RESULT 13
LOCUS
    AW142153
DEFINITION
    EST292377 Normalized rat heart, Banto Soares Rattus sp. cDNA clone
    RGIAG83 5' end similar to peptidylglycine alpha-amidating
    monooxygenase, mRNA sequence.
ACCESSION
    AW142153
VERSION
    AW142153.1 GI:6162030
KEYWORDS
    EST.
SOURCE
    Rattus sp.
    ORGANISM
        Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
        Rattus.
REFERENCE
    1 (bases 1 to 422)
    Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
    Kerlavage,A.R. and Adams,M.D.
TITLE
    Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
    Gene Index
JOURNAL
    Unpublished (1998)
COMMENT
    Other ESTs: EST292376 TC89750
    Contact: Lee, NH
    The Institute for Genomic Research
    9712, Medical Center Drive, Rockville, MD 20850, USA

```

```

tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
For clone availability, additional sequence and expression
information related to this EST please check the TIGR Rat Gene
Index (http://www.tigr.org/tdb/rji/rji.html). To order a clone
contact the ATCC (http://www.atcc.org/atcc.html).
Seq primer: M13 Reverse.
FEATURES
    source
        1..422
            Location/Qualifiers
                /organism="Rattus sp."
                /db_xref="taxon:10118"
                /clone="RGIAG83"
                /clone_lib="Normalized rat heart, Banto Soares"
                /note="Organ: heart; Vector: pT7T3pac; Site_1: EcoRI;
                Site_2: NotI"
BASE COUNT
    85 a 118 c 98 g 121 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 9; Length 422;
Best Local Similarity 89.5%; Pred. No. 4.2e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaccggcgagcgccg 19
    ||||| ||||| |||||
Db 83 CTACCGCGCGAAGGCTGG 65

RESULT 14
LOCUS
    BG298646/c
DEFINITION
    602396847F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4511794 5',
    mRNA sequence.
ACCESSION
    BG298646
VERSION
    BG298646.1 GI:13063508
KEYWORDS
    EST.
SOURCE
    house mouse.
    ORGANISM
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
    1 (bases 1 to 426)
    NTH-MGC http://mgc.nci.nih.gov/.
    National Institutes of Health, Mammalian Gene Collection (MGC)
    Unpublished (1999)
    Contact: Robert Strausberg, Ph.D.
    Email: cgapsb@mail.nih.gov
    Tissue Procurement: The Cepko Laboratory
    cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
    DNA Sequencing by: Incyte Genomics, Inc.
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    http://image.llnl.gov
    Plate: LLAM10395 row: m column: 11
    High quality sequence stop: 327.
FEATURES
    source
        1..426
            Location/Qualifiers
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="IMAGE:4511794"
                /clone_lib="NIH_MGC_94"
                /tissue_type="retina"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 3.3 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC library."
BASE COUNT
    89 a 127 c 129 g 81 t
ORIGIN

```

Query Match 79.0%; Score 15.8; DB 10; Length 426;
 Best Local Similarity 89.5%; Pred. No. 4.2e-03;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccggcggaaggccgg 19
 ||||| | | | | | | |
 Db 359 CTACCGGCGCAAGGCGG 341

RESULT 15

BM376491
 LOCUS
 DEFINITION EBem05_SQ002_G02_R IGF Barley EBem05 library linear EST 10-JAN-2002
 clone EBem05_SQ002_G02 5', mRNA sequence.

ACCESSION BM376491
 VERSION
 KEYWORDS EST.
 SOURCE BM376491.1 GI:18119881

ORGANISM

barley.
 Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.

REFERENCE

AUTHORS Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
 Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.

TITLE

JOURNAL Development of Barley Transcriptome Resources

COMMENT

Contact: Waugh R

Unit of Genomics

Scottish Crop Research Institute

Invergowrie, Dundee, DD2 5DA, Scotland, UK

Tel: 00 44 1382 562731

Fax: 00 44 1382 562426

Email: rwaugh@scri.sari.ac.uk

All sequence has a Phred quality score of 20 or over

Seq primer: M13 reverse.

FEATURES

Location/Qualifiers

1..438

/organism="Hordeum vulgare"

/cultivar="Optic"

/db_xref="taxon:4513"

/clone="EBem05_SQ002_G02"

/clone_lib="IGF Barley EBem05 library"

/tissue_type="Embryo"

/dev_stage="14 days post anthesis"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;

Non-normalised library, directionally cloned into pSPORT1.

Derived from embryos dissected from developing grains (14

days post anthesis) in glasshouse grown barley plants.

Developed as part of the barley transcriptome resources of

BSRC/SEERAD funded cereal IGF (Investigating Gene

Function) project."

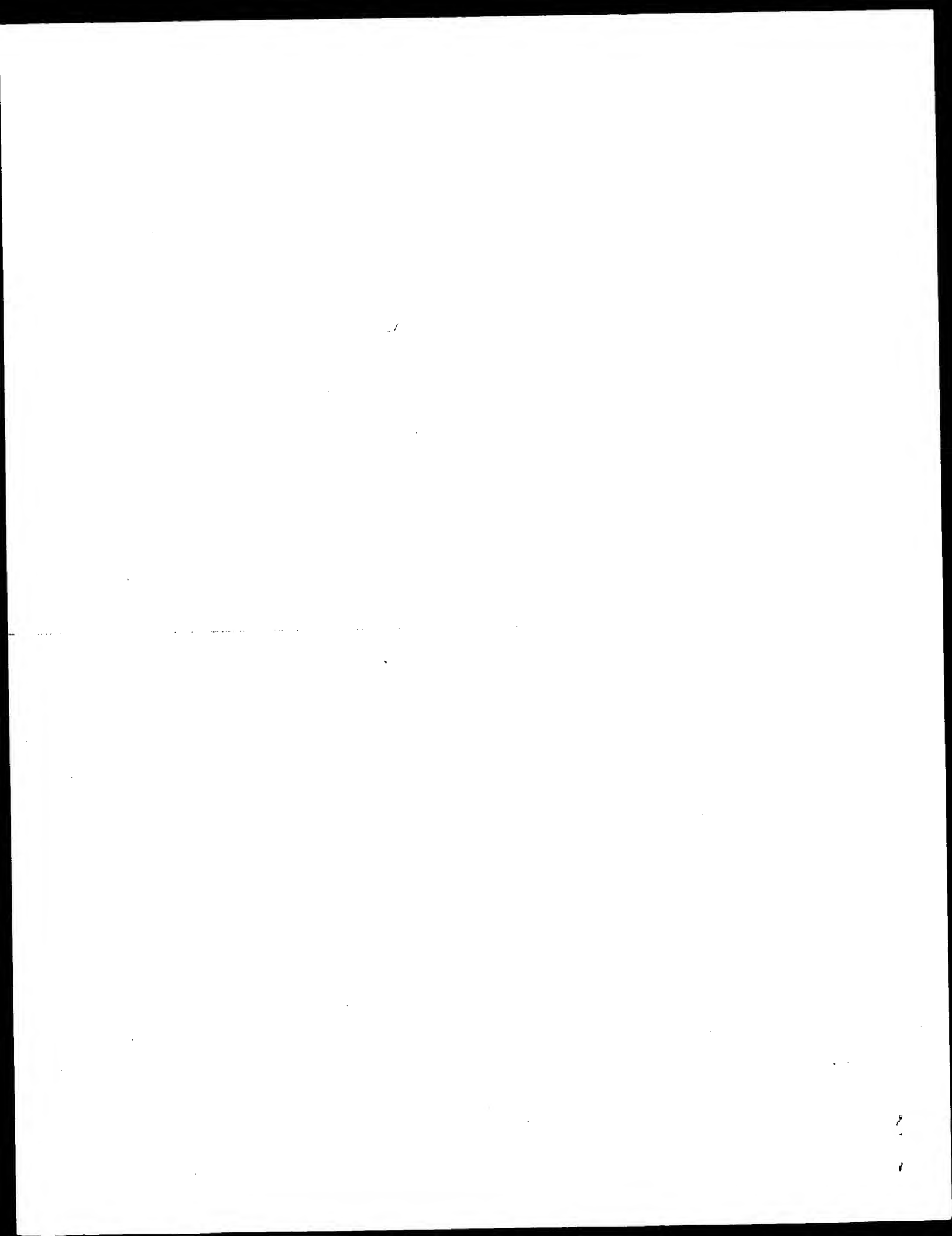
BASE COUNT 87 a 170 c 138 g 43 t

ORIGIN

Query Match 79.0%; Score 15.8; DB 10; Length 438;
 Best Local Similarity 89.5%; Pred. No. 4.2e-03;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctaccggcggaaggccgg 19
 ||||| | | | | | | |
 Db 361 CTACCGGCGGAATGCGG 379

Search completed: June 19, 2002, 14:04:12
 Job time: 7510 sec




```

Matches 691; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atcgagagccagcgacgacgcacacatctctggtatcgagacgagcggagagac 60
Db 601 ATCGAGAGCCAGCGGACGATCGACATCTCTGATATCGGACGAGCGGAGGAGAC 660
Qy 61 ggaagtcctgcttgcacacatctcgactcaagcccgactggtctcacatgtggt 120
Db 661 GGAAGTCCTGCTTGCACATCTCGACTCAAGCCGACTGCTTCTACATGTGGT 720
Qy 121 ggaacacagagcgatgttaccagtacatcgagacacacacacacacacacac 180
Db 721 GGAACACAGGAGCGATGTACAGTACATCGAGACACACACACACACACACAC 780
Qy 181 gatgtaccacagtgtaattagatatttaattatgacctgttagatgtttaaac 240
Db 781 GATGTACCCAGGTGTAATTAGATATTAAATTATGCCCTGTGAGATGTTTAAAC 840
Qy 241 agggcattcagtcgacacacacacacacacacacacacacacacacacacac 300
Db 841 AGGGCATTCAGTTCGACAAATACGACACCCCTAGTTATCTTGGTTCGACCAT 900
Qy 301 gtactcgtattgccaatgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 360
Db 901 GTACTCGTATTGCGCAATGCTCGCTGCTGATTATTGAAATCAGCAGGACAG 960
Qy 361 ctgtggaattattaaagtatgtctatcaataacacacacacacacacacacac 420
Db 961 CTGTGGAATATTAAAGTATGTCTATCTAAATACACCAATACCCGCCGCCG 1020
Qy 421 tatcgtttacattcttaataatctgccccagcgccgaagcctggaggtgctacc 480
Db 1021 TATCGTTTACATCTTATGAATATCTGCCAGCGCGGAGGCTGGAGGTGCT 1080
Qy 481 cgaagccgggacacacacacacacacacacacacacacacacacacacacac 540
Db 1081 CGAAGCCGGGACCAATATGAATCAGTATGAGGCGGCGCCCAATATAAAT 1140
Qy 541 ggaataagacgaatctgttacttctgtcagtcgacgcacacacacacacacac 600
Db 1141 GGATAAGACGAATCTGTTACTTTGCTTGCAGTCGACGACCAACCTTTCC 1200
Qy 601 tcaggtgagtcgctgctgagagagagagagagagagagagagagagagagag 660
Db 1201 TCCAGGTGATGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy 661 ttgcgcgacgagcgtgagtgatctgg 691
Db 1261 TTTCGCGATCGACGCGCTGAGTTGATCTGG 1291

```

```

RESULT 2
CFDCG
LOCUS
DEFINITION
ACCESSION M29963
VERSION M29963.1 GI:323306
KEYWORDS circular; complete genome.
SOURCE Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.
ORGANISM Coconut foliar decay virus
VIRUSES: ssDNA viruses; Nanovirus.
REFERENCE
1 (bases 1 to 1291)
AUTHORS Rohde,W., Randles,J.W., Langridge,P. and Hanold,D.
TITLE Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus
JOURNAL Virology 176, 648-651 (1990)
MEDLINE 90266484
COMMENT Draft entry and printed sequence for [1] kindly submitted by W.Rhode, 15-MAR-1989, for release after publication.
FEATURES
source
1..1291
/organism="Coconut foliar decay virus"
/db_xref="taxon:12474"

```

```

40..70
/note="stem-loop structure"
103..975
/note="ORF 1"
/codon_start=1
/protein_id="AAA42894.1"
/db_xref="GI:323307"
/translation="MGSSIRWCTFLNTEEEAANVVRRIESNLVIAIVGDEVAIS
TGORHOGFIHLKTRRRLQGLKTLVNDRIHLEPTRGSDQNRDYSCKERYVLLHGV
TPGVKRPRLAQRFAEPEDEURLEDPGYRRCVVHGASVETRWAAENPFPPFHHNQ
LELSAIGEPADDDRTILWICGRDGDGKSFVKYLGLKPDWFTTCGGTRKDVLYQIE
DKRLNILDVPRCNLEYLNALLECCKNRAFSSDKYEPFLSYLGFHDHVLVFNWLPD
YKISRDRKRLWNI"
314..775
/note="ORF 2"
/codon_start=1
/protein_id="AAA42895.1"
/db_xref="GI:323308"
/translation="MTGFTWSPVPTNRIETTVRRNGCFSTESRLVLESKGHDWPN
DLRLMNSAMKTQADTADALITELRWNGQDGPLKIRSHFHTITGLSKCCLRSRQT
TAQSSGYADETEETGSPCLPNISDSSTGSHVVEPERTYCTISRTQNEI"
complement(422..568)
/note="ORF 6"
/codon_start=1
/protein_id="AAA42896.1"
/db_xref="GI:323309"
/translation="MEMGTDFORPILSIPPKLRVORIFGIRLGLPGGVHQPQIVGP
IVAF"
639..797
/note="ORF 3"
/codon_start=1
/protein_id="AAA42897.1"
/db_xref="GI:323310"
/translation="MRTRRRREVRCQISRTQARLVLMHWNQKGRIVPVHRGPKTK
FNPRCTQV"
complement(823..987)
/note="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="MTHTLNIPQFYSVPADFQIIRQDYGKYEYMHVPEKITKGFV
RTECPVLNTF"
1098..1286
/note="ORF 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRVMGGPTIKDSIWRINLLCLOCTQPLSTPIONSSLEKK
AASLYLPSICFCAIGRLS"
BASE COUNT 336 a 323 c 300 t
ORIGIN

```

```

Query Match 88.5%; Score 691; DB 14; Length 1291;
Best Local Similarity 100.0%; Pred. No. 7.2e-191;
Matches 691; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atcgagagccagcgacgacgcacacatctctggtatcgagacgagcggagagac 60
Db 601 ATCGAGAGCCAGCGGACGATCGACATCTCTGATATCGGACGAGCGGAGGAGAC 660
Qy 61 ggaagtcctgcttgcacacatctcgactcaagcccgactggtctcacatgtggt 120
Db 661 GGAAGTCCTGCTTGCACATCTCGACTCAAGCCGACTGCTTCTACATGTGGT 720
Qy 121 ggaacacagagcgatgttaccagtacatcgagacacacacacacacacacac 180
Db 721 GGAACACAGGAGCGATGTACAGTACATCGAGACACACACACACACACACAC 780
Qy 181 gatgtaccacagtgtaattagatatttaattatgacctgttagatgtttaaac 240
Db 781 GATGTACCCAGGTGTAATTAGATATTAAATTATGCCCTGTGAGATGTTTAAAC 840

```


JOURNAL Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und Pflanzenvirologie, Biologisches Bundesanstalt fuer Land- und Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY

REFERENCE 2 (bases 1 to 1004)

AUTHOR Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.

TITLE Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome

JOURNAL J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)

MEDLINE 99094637

FEATURES
Location/Qualifiers
1..1004 /organism="faba bean necrotic yellows virus"
/virion
/isolate="SV292-88"
/db_xref="taxon:59817"
73..918 /gene="C9"
73..918 /gene="C9"
/gene="C9"
/function="putative replication associated (rep) protein"
/codon_start=1
/product="component 9"
/protein_id="CAA06789.1"
/db_xref="GI:3550533"
/db_xref="SPTREMBL:O91252"
/translation="MSAVNWPTLNFAGEVPLSFDERVOYVMQHERVNHDIQGV I OLKKAKMNTVNKIIGNPFLHKMKSGTSEASAYAQKEESRVAGPMYSQGLKKGSHK RKTMELIKDPENLEPEEPQLQPIRRAMAWSAMDSEARLKAEEGFPYMFYSWQETVLGLLEE IPNDRTIIVYGPNEGKSQFGPKLGKLDLYLPQGTQDMFTIMLMKNPRANVYMD IPNCNSELNYOFMELIKNRTIYSKYEPVGCIINNKITHIVLANVLPDYEKISQDRI KIIVC"

BASE COUNT 352 a 159 c 226 g 267 t

ORIGIN

Query Match 14.7%; Score 114.8; DB 14; Length 1004;
Best Local Similarity 56.7%; Pred. No. 1.9e-22;
Matches 212; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 2 tcggagaccgaggacatcgcaaatcctctggatatgcgacgacgagcggagagacy 61
| | | | |
| | | | |
DB 545 TAGAAGAAGAGCCCCAATGACCGCTACTATTATTGGGTGTACGCCCAAAATGGTATGAAG 604
| | | | |
| | | | |
QY 62 ggaagtccgtgtttgccaaatatctcggaactcaagcccgcactggttctacacatggtg 121
| | | | |
| | | | |
DB 605 GAAATCACAGTTTGTTAAATTCCTGGGATTAATAAAAAAGATTACCTTTATTTACCTGGAG 664
| | | | |
| | | | |
QY 122 gaaccagaagacgatattgtaccagcatcacatcgaggagccccaaaataaatttaacctcg 181
| | | | |
| | | | |
DB 665 GTAAAACCCAACATGATGCATATATGTTAATGTAAAAATCCAAGGCCAAATGTTGTGATGG 724
| | | | |
| | | | |
QY 182 atgtaccagggtgaatttagtgattttaaattatgcctgttgagaatgtgttaaagaaca 241
| | | | |
| | | | |
DB 725 ATATTCTCGTTGTAATCTGNAATTTAAATTACCATAATTATGGAATTAATTAATAATA 784
| | | | |
| | | | |
QY 242 gggcatttcagttcggacaaaatacgaaccccttagttatcttgggttcgaccatgtgcatg 301
| | | | |
| | | | |
DB 785 GAACCATATATAGTTATTAATATGAACCAAGTTGGATGTATTATAATAATAAATAACAAG 844
| | | | |
| | | | |
QY 302 tactgtatttcgcaatgctcctgcctgattattgaaaaatcagcaggagcagacaataaac 361
| | | | |
| | | | |
DB 845 TAATGTATTAGCTAATGATTTCCTGATTGAAAAAATTAGTACAGTAGAATTAAAA 904
| | | | |
| | | | |
QY 362 tgtggaatatattaa 375
| | | | |
DB 905 TAAATTTATTTGTTAA 918

RESULT 5

AR063452 AR063452 linear PAT 29-SEP-1999

LOCUS Sequence 4 from patent US 5846705.

DEFINITION

ACCESSION	AR063452
VERSION	AR063452.1 GI:5992760
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 1096)
TITLE	Wu,R.-Y., You,L.-R. and Soong,T.-S. Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus
JOURNAL	Patent: US 5846705-A 4 08-DEC-1998;
FEATURES	Location/Qualifiers
source	1..1096
BASE COUNT	347 a 231 c 244 g 274 t
ORIGIN	/organism="unknown"
Query Match	14.6%; Score 114.2; DB 6; Length 1096;
Best Local Similarity	57.5%; Pred No. 2.9e-22;
Matches	225; Conservative 0; Mismatches 163; Indels 3; Gaps 1;
QY	1 atcgagagcaggacgatgcacactctctgatatgcggacgagacgaggagac 60
Db	551 ATCGAAGGTGTTCTGATGATCAGAGTATCATCTGGGTATACGGTCCCAACGGAGCGAA 610
QY	61 ggaaagtccgtgtttccaataatctcgactcaaacgccgactgggtctcacacatgtgt 120
Db	611 GGAAAGTCAACCTTCGCAGAGATATCATCTATAAACCCGGATGGGATATATCAACGGT 670
QY	121 ggaaccagaagcagctattgtaccagtcacatcgagagccccaaaaacgaataatcctc 180
Db	671 GGAAAGCGTCGGATATGATGCACATCATACCGATGGATCCTGATTAATCATTTGGATTAT 730
QY	181 gatgtacccagggtgaatttagagttatttaaatatgccctgttagaagtgttaaagaac 240
Db	731 GATATCCCCAGAAGTCATCAGATTATCTGAATATGGCCTTATAGACACAAAATTAGAAT 790
QY	241 agggcattcagttcggacaatacgaacc---cctagttatcttgggttcgcacctgtg 297
Db	791 AGAGTTTTTAATAACAATAACAAATACGAACCATGTGTGATTAGAAAAAGATGGACAAAATGTC 850
QY	298 catgtactcgtatttgccaattgctcgtcctgattatttgaaaatcagcaggacagaata 357
Db	851 CATGTAATTGTTATGGCAAATGTGTGCCCTCAFTATTGTAAATTTTCAGAAGATAGATA 910
QY	358 aaactgtggaattattaaagtatgtgtcatc 388
Db	911 AAAATAATTAATTGTTGAGAAAGGAACATTC 941
RESULT	6
BYTV2	BYTV2
LOCUS	Banana bunchy top virus (BBTV DNA II) V1, V2, C1 and C2 genes,
DEFINITION	complete cds' s.
ACCESSION	L32167
VERSION	L32167.1 GI:520791
KEYWORDS	stem loop.
SOURCE	Banana bunchy top virus DNA.
ORGANISM	Banana bunchy top virus
REFERENCE	Viruses; ssDNA viruses; Nanovirus.
AUTHORS	1 (bases 1 to 1096)
TITLE	Wu,R.-Y., You,L.-R. and Soong,T.-S. Nucleotide sequences of two circular single-stranded DNAs associated with banana bunchy top virus
JOURNAL	Unpublished (1994)
FEATURES	Location/Qualifiers
source	1..1096
	/organism="Banana bunchy top virus"
	/db_xref="taxon:12585"
TATA signal	1..7

Qy	121	ggaaccagaagagcattgtaccagatcatcgagagacccaacagaa---atttaac	177
Db	676	GGGAAGAGGAGAACATACACTCTCTCCTCGCTGGACGAAGGATCTGAGAAGCATATTGTA	735
Qy	178	ctcgatgaccaggtgaatttagagattttaaattatgccctgttagaattgtttaag	237
Db	736	TTTGCATATCTCCTCGCTGTAATCAGGATTAATTAATATATGATGTTATAGAGGCATTAAAG	795
Qy	238	aacaggcgattcagctcggaacaatacgaaccccttagttatcttgggttcgcacctgtg	297
Db	796	CATAGGCTGATAGAGAGTACTAAATATAAACTATTAAAGTTACTTGAATTGAATATATA	855
Qy	298	catgtactcgtatttgcaatgctcctgctgattatttgaacalcagcagggacacaata	357
Db	856	CATGTAATGTCATGGCTAAATTCATGCCAGAAATCTGTAAAACTCCGAAGATAGAATA	915
Qy	358	aaactgtggaattttaagtagtgtgtcatctaaataaac	398
Db	916	AAGATTATTATTGTTAAATAACACGCTATGACATCGTAC	956
RESULT	11		
AB000922			
LOCUS		1000 bp	DNA
DEFINITION		Milk vetch dwarf virus genome segment 3 encoding virus replication-associated protein, complete sequence.	
ACCESSION		AB000922.1	GI:3808181
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
stem_loop			
CDS			
TATA_signal			
polyA_signal			
BASE COUNT	333 a	185 c	238 g
ORIGIN			

Query Match 11.9%; Score 93; DB 14; Length 1004;
Best Local Similarity 55.3%; Pred. No. 4.5e-16;
Matches 202; Conservative 0; Mismatches 160; Indels

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 16:50:59 ; Search time 3798.65 Seconds
(without alignments)
352.997 Million cell updates/sec

Title: US-09-462-955B-1_COPY_211_991

Perfect score: 781

Sequence: 1 atcggagacgcagggacga.....ccccaacctctgtaaacccc 781

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802:*

```

1: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1980.DAT:*
2: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1981.DAT:*
3: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1982.DAT:*
4: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1983.DAT:*
5: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1984.DAT:*
6: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1985.DAT:*
7: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1986.DAT:*
8: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1987.DAT:*
9: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1988.DAT:*
10: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1989.DAT:*
11: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1990.DAT:*
12: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1991.DAT:*
13: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1992.DAT:*
14: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1993.DAT:*
15: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1994.DAT:*
16: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1995.DAT:*
17: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1996.DAT:*
18: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1997.DAT:*
19: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1998.DAT:*
20: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1999.DAT:*
21: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2000.DAT:*
22: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001A.DAT:*
23: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001B.DAT:*
24: /net/abs06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2002.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114.2	14.6	1096	20	AAV71833
2	114.2	14.6	1096	21	AAV38950
3	112.6	14.4	1091	20	AAV71834
4	112.6	14.4	1091	21	AAV38947
5	112.4	14.4	886	20	AAV34687
6	92.6	11.9	1106	20	AAV71832
7	91	11.7	1106	21	AAV38946
8	89	11.4	1017	17	AAV13165
9	88.4	11.3	1022	17	AAV13161
					BBTV DNA II clone
					Banana bunchy top
					BBTV DNA II clone
					Banana bunchy top
					DNA sequence of BB
					BBTV DNA I clone
					Banana bunchy top
					SCSV segment 6. S
					SCSV segment 2. S

10	81.2	10.4	593	20	AAV34686	DNA sequence of BB
11	62.8	8.0	1110	19	AAV24089	Banana bunchy top
12	59.6	7.6	982	18	AAV49405	Banana bunchy top
13	59.6	7.6	1110	19	AAV24086	Banana bunchy top
14	59.6	7.6	1111	19	AAV24077	Banana bunchy top
15	59.6	7.6	1111	19	AAV24084	Banana bunchy top
16	59.6	7.6	1111	19	AAV24087	Banana bunchy top
17	58.6	7.5	1103	19	AAV24091	Banana bunchy top
18	58.6	7.5	1105	19	AAV24093	Banana bunchy top
19	58	7.4	1109	19	AAV24088	Banana bunchy top
20	58	7.4	1111	19	AAV24090	Banana bunchy top
21	57	7.3	1104	19	AAV24092	Banana bunchy top
22	56.4	7.2	1111	19	AAV24085	Banana bunchy top
23	42	5.4	300	20	AAV71831	Subgenomic fragmen
24	42	5.4	300	21	AAV38949	Banana bunchy top
25	36.8	4.7	287	20	AAV71830	Subgenomic fragmen
26	36.8	4.7	287	21	AAV38948	Banana bunchy top
27	36.8	4.7	2885	23	AAV70474	DNA encoding novel
28	35	4.5	1935	23	AAV68521	DNA encoding novel
29	35	4.5	1935	23	AAV72032	DNA encoding novel
30	35	4.5	1935	23	AAV73099	DNA encoding novel
31	34.8	4.5	4344	23	AAV73169	DNA encoding novel
32	34.8	4.5	4447	23	AAV72881	DNA encoding novel
33	34.8	4.5	5584	23	AAV69664	DNA encoding novel
34	34.6	4.4	1915	23	AAV81517	DNA encoding novel
35	34.6	4.4	2162	23	AAV76891	DNA encoding novel
36	34.2	4.4	951	23	AAV68084	DNA encoding novel
37	34.2	4.4	2523	23	AAV81683	DNA encoding novel
38	34	4.4	997	23	AAV74263	DNA encoding novel
39	34	4.4	1599	23	AAV73145	DNA encoding novel
40	34	4.4	1599	23	AAV74254	DNA encoding novel
41	34	4.4	2011	23	AAV77251	DNA encoding novel
42	34	4.4	3489	23	AAV68093	DNA encoding novel
43	34	4.4	4331	23	AAV72876	DNA encoding novel
c 44	33.8	4.3	1182	23	AAV81606	DNA encoding novel
c 45	33.6	4.3	1479	23	AAV92524	DNA encoding novel

ALIGNMENTS

RESULT 1

AAV71833

ID AAV71833 standard; DNA; 1096 BP.

XX AC AAV71833;

XX AC AAV71833;

XX AC AAV71833;

DT 10-FEB-1999 (first entry)

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

XX BBTV DNA II clone (2-17) nucleotide sequence.

161	gatctaccagggtgtaatttagagatttaaattatgcctgttgaagaattgtttaagaac	240
162		
163		
164		
165		
166		
167		
168		
169		
170		
171		
172		
173		
174		
175		
176		
177		
178		
179		
180		
181		
182		
183		
184		
185		
186		
187		
188		
189		
190		
191		
192		
193		
194		
195		
196		
197		
198		
199		
200		
201		
202		
203		
204		
205		
206		
207		
208		
209		
210		
211		
212		
213		
214		
215		
216		
217		
218		
219		
220		
221		
222		
223		
224		
225		
226		
227		
228		
229		
230		
231		
232		
233		
234		
235		
236		
237		
238		
239		
240		
241	aggcattccagttcggacaaatacgaacc---ccttagttatcttgggttcgaccatgtg	297
242		
243		
244		
245		
246		
247		
248		
249		
250		
251		
252		
253		
254		
255		
256		
257		
258		
259		
260		
261		
262		
263		
264		
265		
266		
267		
268		
269		
270		
271		
272		
273		
274		
275		
276		
277		
278		
279		
280		
281		
282		
283		
284		
285		
286		
287		
288		
289		
290		
291		
292		
293		
294		
295		
296		
297		
298		
299		
300		
301		
302		
303		
304		
305		
306		
307		
308		
309		
310		
311		
312		
313		
314		
315		
316		
317		
318		
319		
320		
321		
3		

PI	Soong T, Wu R, You L;
XX	
XX	
DR	WPI; 1999-059037/05.
DR	p-PSDB; AAW87459.
XX	
XX	Nucleic acids having banana buncy top virus component sequences -
PT	used to design primers for use in polymerase chain reaction
PT	detection of the virus
PT	
XX	Claim 1; Fig 11A-B; 27pp; English.
XX	
XX	
CC	This represents the nucleotide sequence of a banana buncy top virus
CC	(BBTV) DNA I (clone 7-4-2) circular single stranded (css) DNA. The
CC	invention provides nucleic acid sequences associated with BBTV that can
CC	be used in a PCR technique for detecting BBTV. The nucleic acid

XX	(CSIR) COMMONWEALTH SCI & IND RES ORG.
PA	(AUSU) UNIV AUSTRALIAN NAT.
PA	
XX	
PI	Boevink PC, Chu PWG, Keese PK, Khan RI, Larkin PJ;
PI	Marshall JS, Surin BP, Taylor WC, Waterhouse PM;
XX	
XX	WPI; 1996-160363/16.
XX	
PT	Circovirus transcription regulatory sequences and related constructs
PT	- useful in plants, esp. leguminous plants, for the modulation of
PT	gene expression
XX	
PS	Claim 4; Page 62; 121pp: English.
XX	
CC	The subtterranean clover stunt virus (SCSV) genome has at least 7
CC	distinct ssDNA components, designated segments 1-7 (AAT13160-66), each
CC	contg. 1 major open reading frame and a non-coding region. Segment
CC	6 is predicted to be a viral replication-associated protein gene.
CC	Genetic constructs useful in the genetic engineering of plants (esp.
CC	legumes) comprise 1 or more heterologous gene(s) operatively linked
CC	to a promoter region, and in some cases also a terminator region,
CC	selected from segments 1-7. The transcription regulators facilitate
CC	expression of foreign genes in plants and also facilitate control of
CC	levels of gene expression in different plant tissue types.
XX	
SQ	Sequence 1017 BP; 312 A; 160 C; 265 G; 280 T; 0 other;

Query Match	11.4%;	Score 89;	DB 17;	Length 1017;
Best Local Similarity	53.8%;	Pred. No. 1.5e-19;		
Matches 207;	Conservative 0;	Mismatches 175;	Indels 3;	Gaps 1;

Qy	2	tcgagagccgcggacgatcgcacaatctcttgatatcggaacgacgaggagacy	61
Db	529	tagaggaaaccagattatagaacgataactcgttgggtgatgacctcgtgtaatgaaag	588
Qy	62	ggaagtccgtgtttgccaaaatctcggactcaagcccactgtttcacacatgtggtg	121
Db	589	gcaaatctacatttgcagacatctgtcatigaaagatggttggggttatctcctggag	648
Qy	122	gaaccggaagagcgtattgtaccagtcacatcgaggaccacaaacgaaatttaactctcg	181
Db	649	gaaagacacaaagatgatgcatcttgtgaactgtgagccctaagaatacaattgggtatttg	708
Qy	182	atgtaccacaggtgtaatttagagattttaaaattgcctgttagaatgtgttaaagaaca	241
Db	709	acataccacaggtagttccagatgtggaattatggtgttaatagaacaggtttaagaata	768
Qy	242	gggcattcagttcggacaaatacgaacc---ccttagttatcttggttcgaccattg	298
Db	769	gggtaatggtgaatactagtagccatgtgttaatcgvggatgaataatcctctgttc	828
Qy	299	atgtactcgtatttgcgaatgtcctgcctgattttgaaaaatcagcagggcagacaataa	358
Db	829	atgcaatgtgtttgaaatgtactccccagatttgggaaataaaggaagtagaataa	888
Qy	359	aactgtggaataatttaaagtatgtg	383
Db	889	aattaactcgtgtgaaactctg	913

RESULT	9
AAT13161	
ID	AAT13161 standard; DNA; 1022 BP.
XX	
AC	AAT13161;
XX	
DT	23-MAY-1996 (first entry)
XX	
DE	SCSV segment 2.
XX	
KW	SCSV; promoter; transcription; transgenic plant; legume;

Db 790 aggtattattaaattcatggtttatttagagggaatttaagaalggaaataattcaaaagcgggg 849
Qy 260 aatacgaaccccttagttatcttgggttcgaccatgtgcatgtactcgatatttgccaatg 319
Db 850 aatacgaacccgttttgaagatag--tagaataatgtcgaagtcattgtgaatggctaact 906
Qy 320 tcctgcctgattattttgaaaaatcagcaggggacagagaataaaactg 363
Db 907 tccttcggaagggaaggaatctttctctgaagatcgaataaagt 950

Search completed: June 19, 2002, 16:51:06
Job time: 17524 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:49:00 ; Search time 247.21 Seconds
(without alignments)
776.019 Million cell updates/sec

Title: US-09-462-955B-1_COPY_211_991
Perfect score: 781
Sequence: 1 atcggagagccagcgagca.....ccccacctctgttaacccc 781

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```

Database :
Issued_Patents_NA:*
1: /cgm2_6/pdata1/ina/5A_COMB.seq:*
2: /cgm2_6/pdata1/ina/5B_COMB.seq:*
3: /cgm2_6/pdata1/ina/6A_COMB.seq:*
4: /cgm2_6/pdata1/ina/6B_COMB.seq:*
5: /cgm2_6/pdata1/ina/PCTUS_COMB.seq:*
6: /cgm2_6/pdata1/ina/backfiles1.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	691	88.5	1291	4	US-09-462-975-1	Sequence 1, Appli
2	114.2	14.6	1096	2	US-08-418-071-4	Sequence 4, Appli
3	112.6	14.4	1091	2	US-08-418-071-5	Sequence 5, Appli
4	92.6	11.9	1106	2	US-08-418-071-3	Sequence 3, Appli
5	89	11.4	1017	4	US-08-793-634B-6	Sequence 6, Appli
6	88.4	11.3	1022	4	US-08-793-634B-2	Sequence 2, Appli
7	62.8	8.0	1110	1	US-08-202-186-14	Sequence 14, Appli
8	59.6	7.6	982	3	US-08-973-068-28	Sequence 28, Appli
9	59.6	7.6	1110	1	US-08-202-186-11	Sequence 11, Appli
10	59.6	7.6	1111	1	US-08-202-186-9	Sequence 9, Appli
11	59.6	7.6	1111	1	US-08-202-186-12	Sequence 12, Appli
12	59.6	7.6	1111	1	US-08-202-186-24	Sequence 24, Appli
13	58.6	7.5	1103	1	US-08-202-186-16	Sequence 16, Appli
14	58.6	7.5	1105	1	US-08-202-186-18	Sequence 18, Appli
15	58	7.4	1109	1	US-08-202-186-13	Sequence 13, Appli
16	58	7.4	1111	1	US-08-202-186-15	Sequence 15, Appli
17	57	7.3	1104	1	US-08-202-186-17	Sequence 17, Appli
18	56.4	7.2	1111	1	US-08-202-186-10	Sequence 10, Appli
19	42	5.4	300	2	US-08-418-071-2	Sequence 2, Appli
20	36.8	4.7	287	2	US-08-418-071-1	Sequence 1, Appli
c 21	32.6	4.2	5055	4	US-09-242-632A-13	Sequence 13, Appli
c 22	32.2	4.1	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c 23	31	4.0	31	4	US-09-462-975-2	Sequence 2, Appli
c 24	30	3.8	265	3	US-08-289-222E-5	Sequence 5, Appli
25	30	3.8	265	4	US-09-034-526B-5	Sequence 5, Appli
26	30	3.8	885	4	US-08-858-207A-6	Sequence 6, Appli
c 27	30	3.8	2088	4	US-09-351-414-3	Sequence 3, Appli

ALIGNMENTS

```

RESULT      1
US-09-462-975-1
; Sequence 1, Application US/09462975
; Patent No. 6303345
; GENERAL INFORMATION:
; APPLICANT: Rohde, Wolfgang
; APPLICANT: Becker, Dieter
; APPLICANT: Randles, John W.
; APPLICANT: Hehn, Alain
; APPLICANT: Salamini, Francesco
; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER
; FILE REFERENCE: 23232.0003U1
; CURRENT APPLICATION NUMBER: US/09462,975
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/EP98/04345
; PRIOR FILING DATE: 1998-07-13
; PRIOR APPLICATION NUMBER: 19730502.4
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1291
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence
; OTHER INFORMATION: construct
US-09-462-975-1

```

Query Match	88.5%;	Score 691;	DB 4;	length 1291;
Best Local Similarity	100.0%;	Pred. No. 7e-232;		
Matches 691;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	atcggagagccagcgagcatcgccacaatctctggaatcgcgagcagcgagagagac	60	
Db	601	atcggagagccagcgagcatcgccacaatctctggaatcgcgagcagcgagagagac	660	
QY	61	gggaagtcggtgtttgccaataatctcggaactcaagccgactggtttcacacatgtgt	120	
Db	661	gggaagtcggtgtttgccaataatctcggaactcaagccgactggtttcacacatgtgt	720	
QY	121	ggaaccagaaaggagcgtattgtaccagtcacatcgaggaccgaatttaatctctc	180	
Db	721	ggaaccagaaaggagcgtattgtaccagtcacatcgaggaccgaatttaatctctc	780	
QY	181	gatgtacccaagtgtaatttagagcatttaaatatgcctgttagaatgtgttaagaac	240	
Db	781	gatgtacccaagtgtaatttagagcatttaaatatgcctgttagaatgtgttaagaac	840	

Qy	241	aggcattcagttcggacaaatacgaaccccttagttatcttgggttcgaccatgtgc	300
Db	841	aggcattcagttcggacaaatacgaaccccttagttatcttgggttcgaccatgtgc	900
Qy	301	gtactcgtatttgcgaagtctcgtctatttggaaatcacgaggacagaaataaaa	360
Db	901	gtactcgtatttgcgaagtctcgtctatttggaaatcacgaggacagaaataaaa	960
Qy	361	ctgtggaattttaagaatgtgtcatctaaattacaccaataccgcccgccacgcgc	420
Db	961	ctgtggaattttaagaatgtgtcatctaaattacaccaataccgcccgccacgcgc	1020
Qy	421	talcgtttacatcttatgaatactctgccaggccgaagcctggagggtgtctaccgcgc	480
Db	1021	tatcgtttacatcttatgaatactctgccaggccgaagcctggagggtgtctaccgcgc	1080
Qy	481	cgaagcccggaacaaatagaatcgagttatggcgggccacacataaaagattccattt	540
Db	1081	cgaagcccggaacaaatagaatcgagttatggcgggccacacataaaagattccattt	1140
Qy	541	ggataagaacgaatctgttacttcttgcaftgcacgcacacactttccacgtccacaa	600
Db	1141	ggataagaacgaatctgttacttcttgcaftgcacgcacacactttccacgtccacaa	1200
Qy	601	tcacgttgagtagcttgcctggagagaagaacgcgcacgtctatctacgttccatttgtt	660
Db	1201	tcaagftgagtagcttgcctggagagaagaacgcgcacgtctatctacgttccatttgtt	1260
Qy	661	tttgcgcgatcggacgcgtgagttgatctgg	691
Db	1261	tttgcgcgatcggacgcgtgagttgatctgg	1291

RESULT 2
US-08-418-071-4

US 06-416-011-4
; Sequence 4, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR SSDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; TITLE OF INVENTION: METHOD FOR DETECTING

```

:
:
: TOPOLOGY: circular
: MOLECULE TYPE: genomic DNA
: DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"
:
: HYPOTHETICAL: NO
:
: ANTI-SENSE: NO
:
: ORIGINAL SOURCE:
:
: ORGANISM: Banana Bunchy Top Virus (BBTV)
:
: US-08-418-071-4
:

```

Query Match	14.6%;	Score 114.2;	DB 2;	Length 1096;
Best Local Similarity	57.5%;	Pred. No. 6.6e-30;		
Matches	225;	Conservative 0;	Mismatches 163;	Indels 3; Gaps
Qy	1	atcggaagcgcagcgacgatccacaatcctctgatatcgcgacgacgagcgagac	60	
Db	551	ATCGAAGGTGTTCTGATGATCGAAGTATCATCTGGGTATACGTCCTCCACGGAGCGAA	610	
Qy	61	gggaagtcctgtgttgccaaatctctggactcaagccgcgactggtttctacacatgtggt	120	
Db	611	GGAAAGTCACCTTCGCAAGATATCATTCATTAAACCCGGATGGGATATATCAACGGT	670	
Qy	121	ggaaccgaaaggaagcgtatgtaccagtaacatcgagaccacaaagaaatttaaccctc	180	
Db	671	GGAAAGACGTCCGATATGATGCATCATCAACGATGATCTCATTAATCTGGTATT	730	
Qy	181	gatgtaccagtggttaatttagagtatttaaatatgccctgttagaagtgttaagaac	240	
Db	731	GATATCCCGAAGTCATTCAGATTATCTGAATTATGCGCTTTATAGAACAAATTAAGAAT	790	
Qy	241	agggcatttcagttcgggacaaatacgaacc---ccttagttatctttgggttcgaccatgtg	297	
Db	791	AGAGTTTAAATAAATAACAAATAGCAACCATGTGTGATTAGAAAAGATGGACAAATGTC	850	
Qy	298	catgtactcgtatttgccaaatgctccctcgattattgaaatcagcaggggacagaata	357	
Db	851	CATGTAATGTGTATGGCAAAATGTGTCGCTGATTATTGTTAAATTTTCAGAAGTAGAATA	910	
Qy	358	aaactgtggaatatattaaagtatgtgtcatc	388	
Db	911	AAAAATAATTAATTTGTGAGAAAGGAAACTTC	941	

RESULT 3
US-08-418-071-5
; Sequence 5, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETE
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.

US-081-793-634B-6
; Sequence 6, Application US/08793634B
; Patent No. 6211431
; GENERAL INFORMATION:
; APPLICANT: Boevink, Petra C.
; APPLICANT: Surin, Brian P.
; APPLICANT: Keese, Paul K.
; APPLICANT: Chu, Paul W.G.
; APPLICANT: Waterhouse, Peter M.

```

RESULT      4
US-08-418-071-3
; Sequence 3, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTING
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:

```

2	tcggagagcagcgacgatcgacaactcctcggatcgtcgacgagcagcgaggaacg	61
Dy		
529	TAGAGGAAGAACACGATATTAGACGATAAFTCGGTGTATGGACCTGCTGTAATGAAG	588
Db		
62	ggaagtcctgtgttgccaaatatctcggaactcaagccgactggttctacacatgtgtg	121
Qy		
589	GCAAACTTACATTTGCACACATCTGTCATTGAAGATGTTGGGCTTATCTGCCTCGAG	648
Ddb		
122	gaaccgaaagagcgattgtaccagtcacatcgaggagcccaaacgaaatttaactcctcg	181
Qy		
649	GAAGACACAAGATATGATGCATCTTGTACTGCTGAGCCTAAGAATAATTTGGGTATTGT	708
Ddb		
182	atgtaccaggtgtgaatttagagttattaaattagtcctgttagaattgtttaagaaca	241
Qy		
709	ACATACCCAGAGTTAGTTTCAGAGTATGTGAATTTATGTGTAATGAACACAGGTTAAGAATA	768
Ddb		
242	gggcatttcagttcggacaaatacgaacc---cctagttatcttggttcgcaccatgtgc	298
Qy		
769	GGGTAATGGGTGAATACTAAGTATGAGCCATGTGTAATCGGGGATGATTAATCATCTCTGTC	828
Ddb		
299	atgtactcgtatttgccaatcctgcgcctgattttgaaataacagcagcgacagcaataa	358
Qy		
829	ATGTAATTGTGTTTGCAATGTACTCCAGATTGGGGAAATTAAGTGAAGATAGAATAA	888
Ddb		
359	aactgtggaatatattaaagtatgtg	383
Qy		
889	ATTAATATTCGTTGTGAAAACTCTG	913
Ddb		

1


```

RESULT 10
US-08-202-186-9
; Sequence 9, Application US/08202186
; Patent No. 5756708
;
; GENERAL INFORMATION:
;
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
;

```

```

RESULT 11
US-08-202-186-12
; Sequence 12, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner

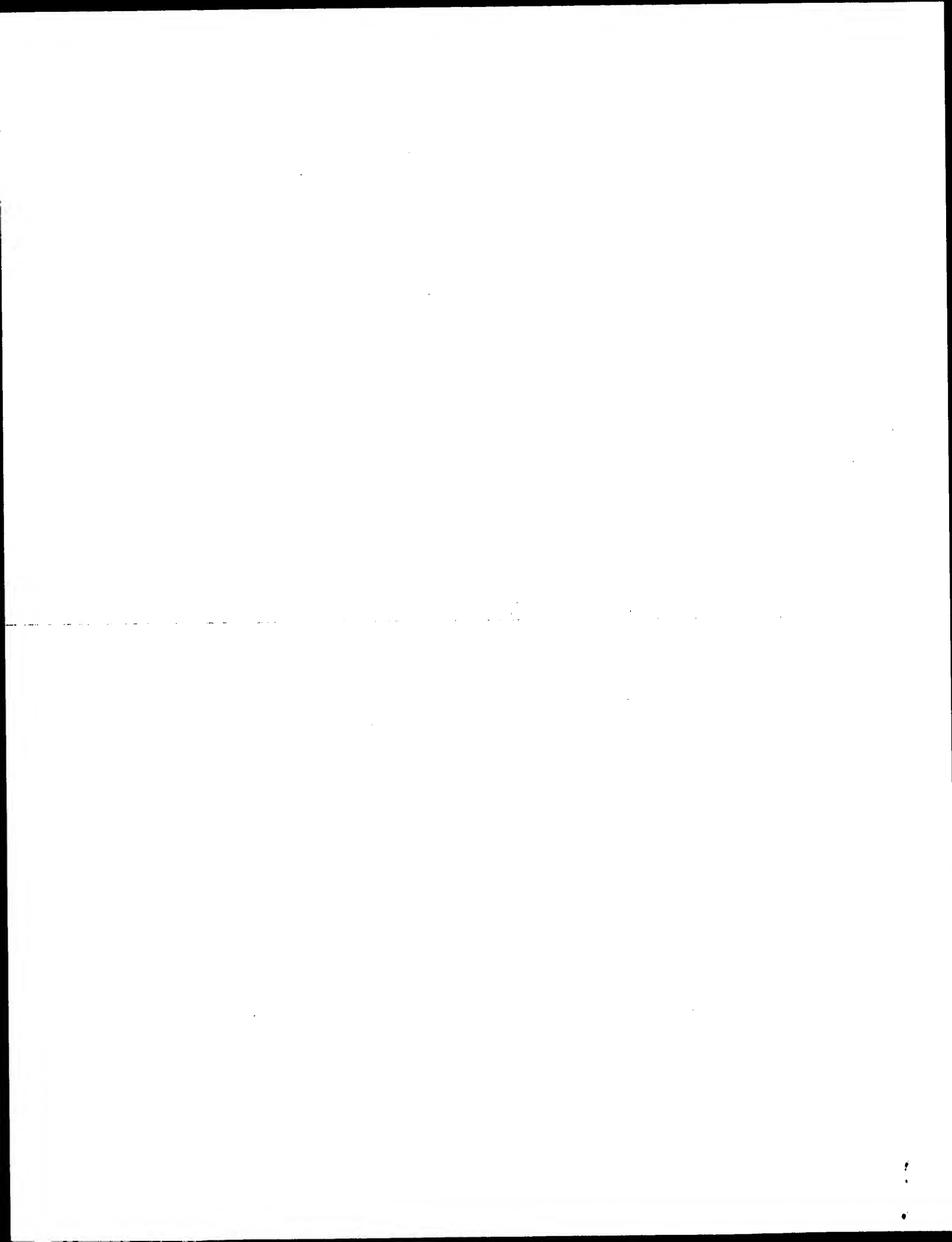
```


RESULT 15
US-08-202186-13
; Sequence 13, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.

```

RESULT 14
US-08-202-186-18
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Foley & Lardner
;

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 14:04:12 ; Search time 7489.97 Seconds
(without alignments)
1407.365 Million cell updates/sec

Title: US-09-462-955B-1_COPY_211_991
Perfect score: 781
Sequence: 1 atcgagagccgagcgagca.....ccccaaacctctgctaacc 781

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	41	5.2	745	9 AV908194	AV908194 AV908194
C 2	40	5.1	512	12 AQ121628	AQ121628 HS_3089_A
C 3	38	4.9	2275	9 AF034173	AF034173 AF034173
4	37.8	4.8	706	9 AL669587	AL669587 AL669587
5	36.8	4.7	498	12 AQ615572	AQ615572 HS_5144_B
6	36.8	4.7	518	9 AW818171	AW818171 CM1-ST027
7	36.4	4.7	787	12 CNS010B7	AL098749 Drosophil
C 8	36	4.6	507	12 A2444169	A2444169 LM0239F20
C 9	36	4.6	929	12 BH137680	BH137680 ENFPC13TR
C 10	35.6	4.6	318	12 AQ089903	AQ089903 HS_3001_A
11	35.6	4.6	1101	12 CNS016H0	AL106734 Drosophil
12	35.4	4.5	460	9 AI942816	AI942816 fc67b05.x
13	35.4	4.5	488	12 CNS0000X	AL085215 Arabidops
C 14	35.2	4.5	345	12 A2401897	A2401897 LM0168N22
15	35.2	4.5	848	12 AQ074604	AQ074604 HS_2277_A
16	35	4.5	524	12 BH175883	BH175883 004_P_17-
17	35	4.5	524	12 CNS071RP	AL612823 T3 end of

18	34.6	4.4	487	12 AQ418525	AQ418525 RPCI-11-2
19	34.6	4.4	555	12 AQ417598	AQ417598 RPCI-11-2
20	34.6	4.4	665	9 BB633121	BB633121 BB633121
C 21	34.6	4.4	689	12 AG167308	AG167308 Pan trogl
C 22	34.4	4.4	594	12 AZ397924	AZ397924 LM0163C12
C 23	34.4	4.4	898	9 AL667528	AL667528 AL667528
24	34.4	4.4	905	10 B1957755	B1957755 HVSMEN001
C 25	34	4.4	442	12 AQ228962	AQ228962 HS_2013_B
26	34	4.4	443	12 AQ010299	AQ010299 HS_2172_B
C 27	34	4.4	1101	12 CNS00DBS	AL067114 Drosophil
C 28	33.8	4.3	410	12 AQ080176	AQ080176 HS_3249_A
C 29	33.8	4.3	604	12 AG019820	AG019820 Homo sapi
C 30	33.8	4.3	673	12 AG167164	AG167164 Pan trogl
C 31	33.8	4.3	749	12 AZ331069	AZ331069 LM0056P10
C 32	33.8	4.3	772	12 AQ488982	AQ488982 RPCI-11-2
C 33	33.8	4.3	865	12 BH135588	BH135588 ENTOC47TF
C 34	33.8	4.3	873	12 CNS012PA	AL101848 Drosophil
35	33.8	4.3	894	12 CNS032CO	AL267441 Tetradon
C 36	33.8	4.3	1174	10 BE779656	BE779656 601464455
C 37	33.6	4.3	326	12 AQ008775	AQ008775 RPCI11-22
38	33.6	4.3	403	10 BG006561	BG006561 QV4-GN014
C 39	33.6	4.3	418	12 AQ632834	AQ632834 RPCI-11-4
40	33.6	4.3	434	12 AQ728714	AQ728714 HS_5462_B
C 41	33.6	4.3	605	12 AQ481591	AQ481591 RPCI-11-2
C 42	33.6	4.3	658	12 AQ077737	AQ077737 CIT-HSP-2
C 43	33.6	4.3	974	10 BG787937	BG787937 SEAU0C007
44	33.4	4.3	283	9 AI065300	AI065300 TENU2187
C 45	33.4	4.3	348	12 AQ348508	AQ348508 RPCI11-11

ALIGNMENTS

RESULT 1

AV908194/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AV908194 745 bp mRNA linear EST 09-NOV-2001
AV908194 Nori Satoh unpublished cDNA library, young adult Ciona
intestinalis cDNA clone rciad75n04 3', mRNA sequence.
AV908194
EST.
Ciona intestinalis.
Ciona intestinalis.
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cloniidae; Ciona.
1 (bases 1 to 745)
Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
Expressed genes in Ciona intestinalis
Unpublished (2000)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source

Location/Qualifiers

1..745

/organism="Ciona intestinalis"

/db_xref="taxon:7719"

/clone="rciad75n04"

/clone_lib="Nori Satoh unpublished cDNA library, young

adult"

/tissue_type="whole animal"

/dev_stage="young adult"

239 a 121 c 139 g 244 t 2 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 122; Conservative

5.2%; Score 41; DB 9; Length 745;

47.5%; Pred. No. 0.12;

0; Mismatches 135; Indels 0; Gaps 0;

```

Qy 103 tgggtctacacatgtgtggaaccagagaagagctattgtaccagtagtacatcgaggaccca 162
Db 644 TGGTTCACAATGCACAGCGAGATTAAAGCTCAAGTAATCAAGTCCACAGGATCCA 585

Qy 163 aaacgaaatttaactctcgatgtacccagggtgtaatttagagattttaattatgcctg 222
Db 584 AAGTGTGTTCTCATCAAAATATAACTACAAAGTGCATAAAGTGTACAATGTTGCTGCATTT 525

Qy 223 ttagaagtgttaagaacagggaattcagttcggagacaaatcacgaacccttagttatctt 282
Db 524 ATAGTCTTAATGAAGGACTAAGGCATGCTGATCTATTATATATCCACACACTGCAATGACTT 465

Qy 283 gggttcgaccatgtgcattgactgatttggccaaatgtcctgcctgattatttgaaatc 342
Db 464 GTGCTGTACAGTTCCATACCTTGGCCATTGGCAGTGTAGTAATTAATCATATTTAACA 405

Qy 343 agcaggggacagataaa 359
Db 404 AACTTTTAGCAAAAAA 388

RESULT 2
A0121628/c
LOCUS
DEFINITION
  HS_3089_Al_B03_MF CIT Approved Human Genomic Sperm Library D Homo
  sapiens genomic clone Plate=3089 Col=5 Row=C, DNA sequence.
ACCESSION
  A0121628
VERSION
  A0121628.1 GI:3498794
KEYWORDS
  GSS.
SOURCE
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 512)
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
JOURNAL
  99380589
MEDLINE
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington
  401 Queen Anne Avenue North, Seattle, WA 98109, USA
  Tel: (206) 616-3618
  Fax: (206) 616-3887
  Email: jwallace@u.washington.edu
  Sequence Tagged Connector
  Plate: 3089 row: C column: 5
  Class: BAC ends
  High quality sequence stop: 512.
  Location/Qualifiers
    1..512
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="Plate=3089 Col=5 Row=C"
    /clone_lib="CIT Approved Human Genomic Sperm Library D"
    /sex="male"
    /note="Organ: sperm; Vector: pHELOBAC11; BAC Clones In
    E-Coli DH10B"
BASE COUNT 174 a 85 c 115 g 136 t 2 others
ORIGIN

Query Match 5.1%; Score 40; DB 12; Length 512;
Best Local Similarity 59.8%; Pred. No. 0.21;
Matches 67; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 278 atcttggttcgaccatgtcgatgtactgtatttgccaatgtcctgcctgattattga 337
Db 179 ATCCTATTGCCCCCATCTGCTTTACTCCACTTTACCAATGCTCTTTTAAAGTATCCCA 120

```

```

Qy 338 aaatcagcaggagacagaataaaactgtggaatatttaataagtagtattgtctatct 389
Db 119 AACTGAACATAGATATGATCAGACTCAGAGAGATTAAATGAAGACACATTT 68

RESULT 3
AF034173
LOCUS
DEFINITION
  AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
  clone ntcon2 contig, mRNA sequence.
ACCESSION
  AF034173
VERSION
  AF034173.1 GI:2707735
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 2275)
  Tripodis,N. and Ragoussis,J.
  Generation of a transcription map in the region immediately
  centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
  boundary
  Unpublished (1997)
JOURNAL
  Contact: Tripodis, Nikos
  Division of Medical and Molecular Genetics
  Guys Hospital
  7th floor, Guy's Tower, London SE1 9RT, UK
  Email: nikos@nki.ni.
  Location/Qualifiers
    1..2275
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /map="6p21.3"
    /clone="ntcon2 contig"
    /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT 438 a 619 c 470 g 599 t 149 others
ORIGIN

Query Match 4.9%; Score 38; DB 9; Length 2275;
Best Local Similarity 16.7%; Pred. No. 1.9;
Matches 32; Conservative 85; Mismatches 75; Indels 0; Gaps 0;

Qy 352 agataaaactgtggaattatttaagtagtgcataatgcataacacacacacccgcg 411
Db 1469 AAAAAAAMWRYKRRKRRKRTGMYKMYRAMMAMACMACWYWKMRGKKC 1528

Qy 412 ccacgcgctatgctttacatcttatgaatcctgccaggccgaagcctggggagtg 471
Db 1529 WKYKYYKYYTSTYKSWRWYWTYYTYWCCTSMKSASCAMRWGMYGMSRSRSY 1588

Qy 472 ctaccggccgaagccgggaacaatatgaatcgagttatggcgccgcccacataaaag 531
Db 1589 GYWGMSGCGYGMTKRYRYYSWTGWTWTTWYMWKSMTRWTTTWTWTTWTTWTTW 1648

Qy 532 attcatttga 543
Db 1649 WTCWTMRKRG 1660

RESULT 4
AL669587
LOCUS
DEFINITION
  AL669587 706 bp mRNA linear EST 14-JAN-2002
  clone 050ZA01 5', mRNA sequence.
ACCESSION
  AL669587
VERSION
  AL669587.1 GI:18142845
KEYWORDS
  EST.
SOURCE
  Ciona intestinalis.
  Ciona intestinalis
  Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
  Phlebobranchia; Clonidae; Ciona.
REFERENCE
  1 (bases 1 to 706)

```


/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 Kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (gl14732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 205 a 74 c 92 g 136 t
 ORIGIN

Query Match 4.6%; Score 36; DB 12; Length 507;
 Best Local Similarity 58.3%; Pred. No. 3.8;
 Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 294 tggcgtactgctgttgcgaatgctcgtcgtgatttgaataacgagggacag 353
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 364 TATTCAGCACTAGAAATGGGAAGCATTTCTCTATTATTAATGATCAGTAGCCATAT 305
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 354 aataaactggaatattaaagtgtgtcatctaaattacacaa 401
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 304 GATATTGCTGCTGAGGTTGAATCATCTTTTATTATTAATACACAA 257

RESULT 9
 BH137680/c

LOCUS BH137680 929 bp DNA linear GSS 07-AUG-2001
 DEFINITION ENTFC13TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 genomic, DNA sequence.
 ACCESSION BH137680
 VERSION BH137680.1 GI:15096741
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica.
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 929)
 Loftus B., Wang Z., Van Aken S. and Fraser C.
 Determination of clone end sequences from Entamoeba histolytica
 Unpublished (2001)
 HMI:IMSS sheared DNA library (2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 17
 High quality sequence stop: 683.
 Location/Qualifiers
 1. 929
 /organism="Entamoeba histolytica"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOS1; Site_1: Bst I; Constructed at The

FEATURES
 source

Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark, a
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999)."

BASE COUNT 288 a 114 c 135 g 392 t
 ORIGIN

Query Match 4.6%; Score 36; DB 12; Length 929;
 Best Local Similarity 49.5%; Pred. No. 5.2;
 Matches 93; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 215 atgocctgttagaattgtttaagacagggcatttcggacaaatcacgacccctta 274
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 666 ATGCCAAGATATAATAAAGTAATTATATGATAATTCAGTCACTTTACACACAAAGT 607
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 275 gttatcttggttcgaccatgctgctgactgctgatttcccaatgctgctgattatt 334
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 606 CTATTATCCAGGTGGAGCATTTCAAATTAATAGTATTATTAAGGCTATGACATTACAAGA 547
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 335 tgaataatcagcagggacagaaataaaactgtggaattattaaagtatgtcatctaaatt 394
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 546 ACAAATTAATTATGGAGCTTCAAGAGTAGAAGGATATTTGGAATATGATTACCTGAACT 487
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 395 acaccaat 402
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 486 TCAACCAT 479

RESULT 10
 AQ089903/c

LOCUS HS_3001_A2_H04_MR CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=3001 Col=8 Row=O, DNA sequence.
 ACCESSION AQ089903
 VERSION AQ089903.1 GI:3458814
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 318)
 Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
 Kellar A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
 Hood L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589

CONTACT: Mahairas GG, Wallace JC, Hood L
 High throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3001 row: 0 column: 8
 Class: BAC ends
 High quality sequence stop: 318.
 Location/Qualifiers
 1. 318
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3001 Col=8 Row=O"

FEATURES
 source

QY	260	aatacgaaccccttagtctcttggttcgaccatgctgcattgctactctgctattggccaatg. 319
		: : : :
Db	726	DGGCGGACACGYKTADTGATATGRKARSGAARGWKTRGTGTRTARTATWGGDRKAGW 785
		: : : :
QY	320	tctgctgctgattattgaaataacgcaggacagagaataaaactgtggaattattaaagta 379
		: : : :
Db	786	TTGTGKTTRKAWCKTDRAKTGARTGAGGWBAGARAVTGWGATWTDABWSTDMRGAG 845
		: : : :
QY	380	tgtgcatcataaatta 395
		: : : :
Db	846	NRATATWATWTVTWT 861
		: : : :
RESULT	12	
AI942816		
LOCUS	AI942816	460 bp mRNA linear EST 07-JUN-2001
DEFINITION	fc67b05_x1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone	
IMAGE:	3726417 3', mRNA sequence.	
ACCESSION	AI942816	
VERSION	AI942816.1	GI:5707472
KEYWORDS	EST.	
SOURCE	zebrafish.	
ORGANISM	Danio rerio	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;	
	Cyprinidae; Danio.	
	1 (bases 1 to 460)	
REFERENCE	Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy	
AUTHORS	, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, F., Underwood	
	, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,	
	Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,	
	Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.	
	and Wilson, R.	
TITLE	WashU Zebrafish EST Project 1998	
JOURNAL	Unpublished (1998)	
COMMENT	Contact: Stephen L. Johnson	
	Washington University School of Medicine	
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA	
	Tel: 314 286 1800	
	Fax: 314 286 1810	
	Email: zbrafish@watson.wustl.edu	
	cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:	
	Matthew Clark. DNA Sequencing by: Washington University Genome	
	Sequencing Center Clone Distribution: Genome Systems, St. Louis,	
	Missouri (web address: www.genomesystems.com) (email contact:	
	info@genomesystems.com) and Research Genetics, Huntsville, Alabama	
	(web address: www.resgen.com) (email contact: info@resgen.com) and	
	Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:	
	www.rzpd.de)	
FEATURES	Seq primer: T7 ET from Amersham	
	High quality sequence stop: 443.	
	Location/Qualifiers	
source	1..460	
	/organism="Danio rerio"	
	/db_xref="taxon:7955"	
	/clone="IMAGE:3726417"	
	/clone_lib="zebrafish WashU MPIMG EST"	
	/sex="mixed"	
	/tissue.type="26 somite embryos, adult livers, shield	
	stage embryos"	
	/lab_host="Xl1-blue MRF"	
	/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st	
	strand cDNA was primed with a NotI oligo (dp15 primer	
	[5'-pGACTAGTCTAGATCGAGCGGCCCTTTTCTTTTCTTTT3'];	
	double-stranded cDNA was ligated to Sal I adaptors (BRL),	
	digested with NotI and cloned into the NotI and SalI	
	sites of the pSPORT1 vector (BRL). Library was constructed	
	by Matthew Clark (Lehrach lab: ICRF, London and Max Planck	
	Institut fuer Molekulare Genetik, Berlin). cDNAs for EST	
	analysis were selected following oligonucleotide	
	hybridization fingerprinting of arrayed clones from	
	zebrafish late somitogenesis (26 ss), adult liver or	
	embryonic shield stage (5.6 hr) libraries. Fingerprint	

BASE COUNT	77 a	66 c	66 g	108 t	1 others
ORIGIN	/clone_lib="CIT Approved Human Genomic Sperm Library D" /sex="male" /note="organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"				
Query Match	4.6%;	Score 35.6;	DB 12;	Length 318;	
Best Local Similarity	54.6%;	Pred. No. 4.1;			
Matches	71;	Conservative 0;	Mismatches 59;	Indels 0;	Gaps 0;
Qy	232	gttaagaacagggcattcagttcggacaaaatacgaaccccttagttatcttgggttcgac	291		
Db	212	gTCTCGGCAGCGCAATCATCGCAGGAGGAAGAAATAAGGGTATTTAATTAGGAAGAG	153		
Qy	292	catgTgcattactcgtatttcccaatgctcctgcctgatttggaaatcagcagggac	351		
Db	152	GAFTTCAAAATTGTCCTGTTTCAGATGACATGACTGTATATCTAGAAAAACCGCATTTGTC	93		
Qy	352	agaataaaac	361		
Db	92	TCAGTCCAAC	83		
RESULT 11					
CNS016H0	CNS016H0 1101 bp DNA linear GSS 26-JUL-1999				
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC				
DEFINITION	BACN16B20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL106734				
VERSION	AL106734.1 GI:5623558				
KEYWORDS	GSS;				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
AUTHORS	1 (bases 1 to 1101)				
JOURNAL	Genoscope.				
COMMENT	Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (DrosBAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.				
FEATURES	Location/Qualifiers				
source	1..1101				
	/organism="Drosophila melanogaster" /plasmid="pBelOBAC11" /db_xref="taxon:7227" /clone_lib="DrosBAC" /clone="BACN16B20" /note="end : T7"				
BASE COUNT	260 a	123 c	159 g	217 t	342 others
ORIGIN					
Query Match	4.6%;	Score 35.6;	DB 12;	Length 1101;	
Best Local Similarity	30.1%;	Pred. No. 7.6;			
Matches	59;	Conservative 51;	Mismatches 86;	Indels 0;	Gaps 0;
Qy	200	taggtatttaattatgcctcgttagaattgtttaagaacagggcattcgttcgaca	259		
Db	666	TAAATATTAATTAATTAATTAATTTTAAAGAAARTTGWGGGAKADWTTTATNKKGR	725		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:43:27 ; Search time 5904.86 Seconds
(without alignments)
2066.124 Million cell updates/sec

Title: US-09-462-955B-1_COPY_409_991

Perfect score: 583

Sequence: 1 tttagagtatttaattatgc.....ccccaaacctgtgtaacccc 583

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
No.	Score	Match	Length	ID		

1	493	84.6	1291	6	ARI72770	ARI72770 Sequence
2	493	84.6	1291	14	CFDCG	M29963 Coconut fol
3	68.4	11.7	1007	14	FBI32187	AJ132187 faba bean
4	65.2	11.2	1004	14	NYV5966	AJ005966 faba bean
5	59.8	10.3	1111	14	NYV12586	U12586 Banana bunc
6	59.4	10.2	1127	14	BBU12587	U12587 Banana bunc
7	58	9.9	1106	6	AR063451	AR063451 Sequence
8	58	9.9	1106	14	BYTV1	L32166 Banana bunc
9	57.8	9.9	1106	14	U02312	U02312 Banana bunc
10	57.4	9.8	1109	14	AF216221	AF216221 Banana bu
11	52.2	9.0	1091	6	AR063453	AR063453 Sequence
12	52.2	9.0	1096	6	AR063452	AR063452 Sequence
13	52.2	9.0	1096	14	BYTV2	L32167 Banana bunc
14	52	8.9	1095	14	AF216222	AF216222 Banana bu
15	49.6	8.5	1007	14	AB000920	AB000920 Milk vetc
16	46	7.9	1000	14	AB000922	AB000922 Milk vetc
17	45.6	7.8	1017	6	ARI45398	ARI45398 Sequence
18	45.6	7.8	1017	14	SCU16735	U16735 Subterranea
19	45	7.7	1015	14	FBI32185	AJ132185 faba bean
20	41.8	7.2	1343	8	ACY14168	Y14168 Ageratum co
21	41.4	7.1	1367	14	AYE238493	AJ238493 Nanovirus
22	41	7.0	1014	14	NYV5964	AJ005964 faba bean
23	40.2	6.9	151886	9	AC022335	AC022335 Homo sapi
24	40	6.9	26445	9	AC084298	AC084298 Homo sapi
25	40	6.9	125020	9	AF429315	AF429315 Homo sapi
26	39.8	6.8	439	8	AF158545	AF158545 Pinus tae
27	38.6	6.6	1022	14	AB009047	AB009047 Milk vetc
28	38.6	6.6	118595	9	AC002379	AC002379 Human HAC
29	37.4	6.4	1004	14	NYV5968	AJ005968 faba bean
30	37.4	6.4	16473	5	AP002927	AP002927 Polymixia
31	37	6.3	210115	9	AL442127	AL442127 Human DNA
32	36.8	6.3	36652	9	AC005176	AC005176 Homo sapi
33	36.8	6.3	55801	2	AC087476	AC087476 Homo sapi
34	36.8	6.3	102456	9	AP000083	AP000083 Homo sapi
35	36.8	6.3	140770	2	AF252832	AF252832 Homo sapi
36	36.8	6.3	152812	2	AC090536	AC090536 Homo sapi
37	36.8	6.3	153684	2	AC013645	AC013645 Homo sapi
38	36.8	6.3	160556	2	AC013645	AC013645 Homo sapi
39	36.8	6.3	170414	2	AC026019	AC026019 Homo sapi
40	36.8	6.3	172820	2	AC092899	AC092899 Homo sapi
41	36.8	6.3	176730	9	AC055839	AC055839 Homo sapi
42	36.8	6.3	182529	2	AC064832	AC064832 Homo sapi
43	36.8	6.3	187927	9	AC046142	AC046142 Homo sapi
44	36.6	6.3	136232	2	AC024006	AC024006 Homo sapi
45	36.6	6.3	142555	9	AL139214	AL139214 Human DNA

ALIGNMENTS

RESULT	1					
ARI72770						
LOCUS	ARI72770	1291 bp	DNA	linear	PAT 17-DEC-2001	
DEFINITION	Sequence 1 from patent US 6303345.					
ACCESSION	ARI72770					
VERSION	ARI72770.1	GI:17912261				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 1291)					
AUTHORS	Rohde,W., Becker,D., Randles,J.W., Hehn,A. and Salamini,F.					
TITLE	Use of a virus DNA as promoter					
JOURNAL	Patent: US 6303345-A 1 16-OCT-2001;					
FEATURES	Location/Qualifiers					
source	1..1291					
BASE COUNT	336 a	323 c	332 g	300 t		
ORIGIN	/organism="unknown"					

Query Match 84.6%; Score 493; DB 6; Length 1291;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;

Qy 421 tggagaagaacgcgcagcctctatctaccgtccattgttttgcgcgcatcgacgc 480
 Db 1219 TGGAGAAGAAGCGCGAAGCCTCTATCTACCGTCCATTGTTTTCGCCGATCGACGCG 1278
 Qy 481 tgaattgatctgg 493
 Db 1279 TGAGTTGATCTGG 1291

RESULT 3
 FBEL132187
 LOCUS
 DEFINITION faba bean necrotic yellows virus C9-Eg gene, isolate Egyptian EVI-93.
 ACCESSION AJ132187
 VERSION AJ132187.1 GI:4995171
 KEYWORDS C9-Eg gene; rep protein.
 SOURCE faba bean necrotic yellows virus.
 ORGANISM faba bean necrotic yellows virus.
 Viruses; ssDNA viruses; Nanovirus.
 REFERENCE 1 (bases 1 to 1007)
 Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.
 Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
 J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
 99094637
 2 (bases 1 to 1007)
 Timchenko,T., de Kouchkovsky,F., Katul,L., David,C., Vetten,H.J.
 and Gronenborn,B.
 A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants
 J. Virol. 73 (12), 10173-10182 (1999)
 20027244
 REFERENCE 3 (bases 1 to 1007)
 Katul,L.
 Direct Submission
 Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie, Mikrobiol., u. Biol. Sicherheit, Biologische Bundesanstalt f. Land- u. Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY
 Location/Qualifiers
 1..1007
 /organism="faba bean necrotic yellows virus"
 /viroon
 /isolate="Egyptian EVI-93"
 /db_xref="taxon:59817"
 /lab_host="Vicia faba"
 1..39
 stem_loop
 gene 73..1003
 /gene="C9-Eg"
 CDS 73..918
 /gene="C9-Eg"
 /function="replication-associated protein"
 /note="32.53 kDa"
 /codon_start=1
 /evidence=experimental
 /product="Rep protein"
 /protein_id="CAB44027.1"
 /db_xref="GI:4995172"
 /db_xref="SPTREMBL:O9WK2"
 /translation="MSAVNWVFLNAGEVPLVDFERVQVAVWQHERVNHIDHTQGV
 OLKKAKMTVNI IGGNPHLEKMGSIIEASAYAQKEESRVAGPWSYGLLLKGGSHK
 RKIMELIKDPENLEEPQYRRAMASAMDSRKLAEEGFPYTYLSWOETVLGLLEE
 EPNDRTLIWYGPNGNEKSGQFGLGKDYLYLPGGKTQDQTYMLMKNPKANVMD
 IPRCSEYLNQFMELIKNRTIFSYKYPVPGCIINNKTHIVLVANLPDYEKISQDRI
 KIITYC"
 TATA_signal 939..944
 /gene="C9-Eg"
 polyA_signal 998..1003
 /gene="C9-Eg"
 BASE COUNT 353 a 161 c 227 g 266 t
 ORIGIN

Qy 421 tggagaagaacgcgcagcctctatctaccgtccattgttttgcgcgcatcgacgc 480
 Db 1219 TGGAGAAGAAGCGCGAAGCCTCTATCTACCGTCCATTGTTTTCGCCGATCGACGCG 1278
 Qy 481 tgaattgatctgg 493
 Db 1279 TGAGTTGATCTGG 1291

Query Match 11.7%; Score 68.4; DB 14; Length 1007;
 Best Local Similarity 62.1%; Pred. No. 3.7e-09;
 Matches 108; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 4 gagtattaaatgcctgttagaattgttaagaacagggcattcagtcgacaaa 63
 Db 745 GAATATTTAAATCAATTTATGAATTAATAAATAAGAACCACTATTTAGTTATA 804
 Qy 64 tacgaaccccttagttatcttgggttcgaccatgtgcagtactcgatttgcacaatgc 123
 Db 805 TAIGAACCACTGGATGATTAATAAATAAGAACCACTATTTAGTTATA 864
 Qy 124 ctgcctgattattgaaatcagcaggacagacataaaactgtggaattattaa 177
 Db 865 TTGCCTGATTATGAAAAAATAGTCAGCACACAAATTAATAAATTTATGTTAA 918

RESULT 4
 NIV5966
 LOCUS
 DEFINITION faba bean necrotic yellows virus C9 gene.
 ACCESSION AJ005966
 VERSION AJ005966.1 GI:3550532
 KEYWORDS C9 gene; component 9; putative; rep protein; replication associated protein.
 SOURCE faba bean necrotic yellows virus.
 ORGANISM faba bean necrotic yellows virus.
 Viruses; ssDNA viruses; Nanovirus.
 REFERENCE 1 (bases 1 to 1004)
 Katul,L.
 Direct Submission
 Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY
 2 (bases 1 to 1004)
 Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.
 Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
 J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
 99094637
 Location/Qualifiers
 1..1004
 /organism="faba bean necrotic yellows virus"
 /viroon
 /isolate="SV292-88"
 /db_xref="taxon:59817"
 73..918
 /gene="C9"
 73..918
 /gene="C9"
 /function="putative replication associated (rep) protein"
 /codon_start=1
 /product="component 9"
 /protein_id="CAA06789.1"
 /db_xref="GI:3550533"
 /db_xref="SPTREMBL:O91252"
 /translation="MSAVNWVFLNAGEVPLVDFERVQVAVWQHERVNHIDHTQGV
 OLKKAKMTVNI IGGNPHLEKMGSIIEASAYAQKEESRVAGPWSYGLLLKGGSHK
 RKIMELIKDPENLEEPQYRRAMASAMDSRKLAEEGFPYTYLSWOETVLGLLEE
 EPNDRTLIWYGPNGNEKSGQFGLGKDYLYLPGGKTQDQTYMLMKNPKANVMD
 IPRCSEYLNQFMELIKNRTIYSKYEPVPGCIINNKTHIVLVANLPDYEKISQDRI
 KIITYC"
 BASE COUNT 352 a 159 c 226 g 267 t
 ORIGIN

Query Match 11.2%; Score 65.2; DB 14; Length 1004;
 Best Local Similarity 60.9%; Pred. No. 3.2e-08;
 Matches 106; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

[illegible]

RESULT	6	
BBU12587		
LOCUS	BBU12587	1127 bp DNA
DEFINITION	Banana bunchy top virus DNA IV ORF VI and ORF C1 genes, complete cds.	circular VRL 01-FEB-1995
ACCESSION	U12587	
VERSION	U12587.1	GI:642393
KEYWORDS		
SOURCE	Banana bunchy top virus.	
ORGANISM	Banana bunchy top virus	
	Viruses; ssDNA viruses; Nanovirus.	
REFERENCE	1 (bases 1 to 1127)	
AUTHORS	Wu, R.-Y. and You, L.-R.	
TITLE	Nucleotide sequences of DNA III and DNA IV associated with banana bunchy top virus and their relation to other closely related virus	

Unpublished
2 * (bases 1 to 1127)
Wu, R.-Y.
Direct Submission
Submitted (23-JUL-1994) Rey-Yuh Wu, Agricultural Biotechnology
Division, Development Center for Biotechnology, 81 Chang Hsing
Street, Taipei, 10671, Taiwan, ROC
Location/Qualifiers
1..1127
source
/organism="Banana bunchy top virus"
/db_xref="taxon:12585"
/clone="S15 and 4-2"

```

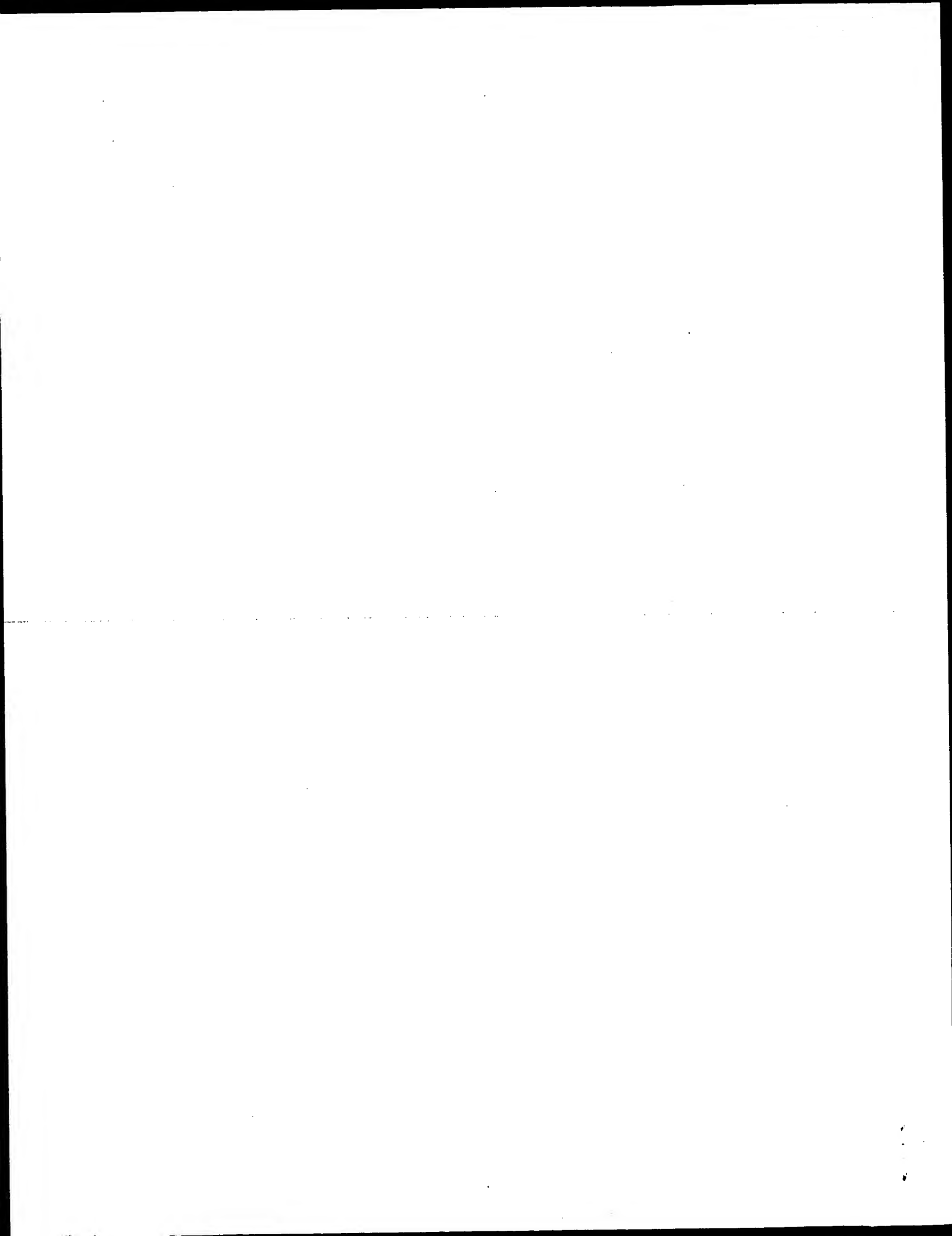
1. .8
15. .49
79. .933
CDS
TATA_signal
stem_loop
/note="ORF V1"
/codon_start=1
/protein_id="AAA61877.1"
/db_xref="GI:642394"
/translation="MSSEKVCFTLNTYSSAAREDFLLALKEELNAVYVGVDEVPSSQ
QKHLQYLSLKKSLKGLKKKYSRAHWRARGSDNKKYCSKETLLLELGFPAAS
GSNRKLLSEMYVSRRPERMRITQPIYHRYTSKRAHFKKEEFVHCLDRPWOIOLTE
IDEEPDDRTITWYQPGNGKESTYAKSLMKKQWVYTRGKGKENILFSTYDVGSGSKH
VFDIPTRCQNDLNTDVTEALKDRVIESTKYKPKLVELNNIHVIVMANPMPFECKLS
DRIKIIVC"
complement(497..643)
/note="ORF C1"
/codon_start=1
/protein_id="AAA61878.1"
/db_xref="GI:642395"
/translation="MTSHMLSPHYHSDHPRRSFDHLVPRQLPPSIESAMVYRGKD
RIPP"

```

	BASE COUNT	353 a	222 c	260 g	292 t	
ORIGIN						
Query Match						
Best Local Similarity			10.2%;	Score 59.4;	DB 14;	Length 1127;
Matches 111;			Conservative	56.3%;	Pred. No. 1.7e-06;	
				0;	Mismatches 86;	Indels 0;
					Gaps	0
Qy	4	gagttattaaattatgccctgttagaattgtttaagaacaggcgcatcagttcgcgacaaa	63			
Db	760	GATTATTAAATTATGATGTTATTAGGCAATTAAAGGATAGGCGTATAGAGAGTACTAAA	819			

REFERENCE	1 (bases 1 to 1095)
AUTHORS	Yeh,H.H., Su,H.J. and Chao,Y.
TITLE	Genome characterization and identification of viral-associated dsDNA component of banana bunchy top virus
JOURNAL	Virology 198, 645-652 (1994)
MEDLINE	94120739
REFERENCE	2 (bases 1 to 1095)
AUTHORS	Chao,Y.
TITLE	Direct Submission
JOURNAL	Submitted (04-OCT-1993) Yu-Chan Chao, Institute of Molecular Biology, 120, Sec. 2., Yen-joe-yuan Rd., Nan-Kang, Taipei, Taiwan 115, Republic of China
FEATURES	Location/Qualifiers
source	1..1095
	/organism="Banana bunchy top virus"
	/specific_host="Musa acuminata Colla"
	/db_xref="taxon:12585"
	/clone="pBTD18"
	/note="component 2 (BBTV-C2)"
CDS	82..390
	/note="orf1"
	/codon_start=1
	/protein_id="AAAL7783.1"
	/db_xref="GI:475434"
	/translation="MSSPSLKKWCFTINYSSAARENFLSLKKEEDVHYAVGVDEVAPA TQGHQLQGYSLSLKMKMPRIIEEVLPCSLGDCQRNRRREFEVLFQRPNPRIYRVSCK"
CDS	419..634
	/note="orf2"
	/codon_start=1
	/protein_id="AAAL7784.1"
	/db_xref="GI:475435"
	/translation="MIARSPDRMKIEQPEIFHRYQSVNKLKKFKEEFVHPCLDRPQWQI QLTEAIDEEPDDESIILWVGYPXGNCK"
BASE COUNT	337 a 217 c 253 g 288 t
ORIGIN	
Query Match	9.9%; Score 57.8; DB 14; Length 1095;
Best Local Similarity	55.8%; Pred. No. 5e-06;
Matches 110; Conservative	0; Mismatches 87; Indels 0; Gaps 0;
Qy	4 gagtattaaattatgccctgtttagaatggtttaagacaggcgattcagtcggacaaa 63
Dd	767 GATTATTAATTATGTATGAATAGAGGCATTAAGAGTAGGGTTATAGAGACTAAA 826
Qy	64 tacgaacccttagtatcttggtgtcgacctgactcgtattgcctaagtgc 123
Dd	827 TACAACCCATAAGATAGTTCATATAGGTAAATACATGTATTCGTCATGCCGAATTC 886
Qy	124 ctgcctgattattgaaaaatcagcaggacagataaaaactgtggaatatttaaatgatg 183
Dd	887 ATGCGTGACTTCTGTAAAAATCTCCGAGATCGAATAAAATCAATTTATTGCTGAATAACA 946
Qy	184 tgtcatctaaattacac 200
Dd	947 CGCTATGACAAATCGTAC 963
RESULT 10	
AF216221	
LOCUS	Banana bunchy top virus satellite SL replication initiation protein 1109 bp DNA linear VRL 08-MAR-2001
DEFINITION	(ORF V1) gene, complete cds.
ACCESSION	AF216221
VERSION	AF216221.1 GI:12004325
KEYWORDS	.
SOURCE	Banana bunchy top virus.
ORGANISM	Banana bunchy top virus Viruses; ssDNA viruses; Nanovirus. 1 (bases 1 to 1109) Horser,C., Harding,R. and Dale,J. Banana bunchy top nanovirus DNA-1 encodes the 'master' replication TITLE

Query Match 9.0%; Score 52.2; DB 14; Length 1096;



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 16:51:06 ; Search time 3798.65 Seconds
(without alignments)
263.504 Million cell updates/sec

Title: US-09-462-955b-1_COPY_409_991
Perfect score: 583
Sequence: 1 tttagtatttaattatgc.....cccccaacctctgtaacccc 583

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802:*

1: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT:*

2: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT:*

3: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT:*

4: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT:*

5: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT:*

6: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT:*

7: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT:*

8: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT:*

9: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT:*

10: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT:*

11: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT:*

12: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:*

13: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT:*

14: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT:*

15: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT:*

16: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT:*

17: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:*

18: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:*

19: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:*

20: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:*

21: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:*

22: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:*

23: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:*

24: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	9.9	1106	20	AAV71832
2	58	9.9	1106	21	AAV38946
3	57.4	9.8	1091	20	AAV34686
4	52.2	9.0	1091	21	AAV71834
5	52.2	9.0	1091	21	AAV38947
6	52.2	9.0	1096	20	AAV71833
7	52.2	9.0	1096	21	AAV38950
8	52	8.9	886	20	AAV34687
9	45.6	7.8	1017	17	AAT13165
					SCSV segment 6. S

10	36.8	6.3	2885	23	AAV70474	DNA encoding novel
11	35	6.0	1935	23	AAV68521	DNA encoding novel
12	35	6.0	1935	23	AAV72032	DNA encoding novel
13	35	6.0	1935	23	AAV73099	DNA encoding novel
14	34.6	5.9	1103	19	AAV24091	Banana bunchy top
15	34.6	5.9	1105	19	AAV24093	Banana bunchy top
16	34.6	5.9	1915	23	AAV81517	DNA encoding novel
17	34.6	5.9	2162	23	AAV76891	DNA encoding novel
18	34.2	5.9	951	23	AAV68084	DNA encoding novel
19	34.2	5.9	2523	23	AAV81683	DNA encoding novel
20	34	5.8	997	23	AAV74263	DNA encoding novel
21	34	5.8	1110	19	AAV24089	Banana bunchy top
22	34	5.8	1111	19	AAV24087	Banana bunchy top
23	34	5.8	1599	23	AAV73145	DNA encoding novel
24	34	5.8	1599	23	AAV74254	DNA encoding novel
25	34	5.8	2011	23	AAV77251	DNA encoding novel
26	34	5.8	3489	23	AAV68093	DNA encoding novel
27	34	5.8	4331	23	AAV72876	DNA encoding novel
28	33.8	5.8	1182	23	AAV81606	DNA encoding novel
29	33.6	5.8	1479	23	AAV92524	DNA encoding novel
30	33.6	5.8	1638	23	AAV68724	DNA encoding novel
31	33.6	5.8	2610	23	AAV83845	DNA encoding novel
32	33.6	5.8	42488	22	AAV66772	Human immune/haema
33	33.2	5.7	1011	23	AAV74571	DNA encoding novel
34	33	5.7	1104	19	AAV24092	Banana bunchy top
35	33	5.7	9345	21	AAV93966	Mouse uromodulin p
36	32.6	5.6	2313	23	AAV74278	DNA encoding novel
37	32.6	5.6	2313	23	AAV79107	DNA encoding novel
38	32.6	5.6	2313	23	AAV92935	DNA encoding novel
39	32.6	5.6	2414	23	AAV84393	DNA encoding novel
40	32.6	5.6	3368	23	AAV68231	DNA encoding novel
41	32.4	5.6	982	18	AAV49405	Banana bunchy top
42	32.4	5.6	1109	19	AAV24088	Banana bunchy top
43	32.4	5.6	1111	19	AAV24077	Banana bunchy top
44	32.4	5.6	1111	19	AAV24084	Banana bunchy top
45	32.4	5.6	2207	23	AAV71746	DNA encoding novel

ALIGNMENTS

RESULT 1
AAV71832
ID AAV71832 standard; DNA; 1106 BP.

XX AC AAV71832;

XX DT 10-FEB-1999 (first entry)

XX DE BBTV DNA I clone (7-4-2) nucleotide sequence.

XX DE Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
XX DE Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.

XX OS Banana bunchy top virus.

XX FH Key Location/Qualifiers

FT TATA_signal 1..8

FT stem_loop /*tag= a

FT CDS /*tag= b

FT /*tag= c

FT /product= "ORF-V2 product"

FT /transl_except= (pos:335..337, aa:Gly)

FT /transl_except= (pos:518..520, aa:Ser)

FT polyA_signal 380..385

FT /*tag= d

FT polyA_signal 901..906

XX /*tag= e

XX US5846705-A.

XX

CC sequence derived from an untranslated portion of any one BBTv components
 CC and is adaptable for promoting transcription of a cloned gene in a plant
 CC cell. The promoter, included in a DNA chimeric vector, is useful for the
 CC expression of a gene in a plant cell. The inclusion of the polyubiquitin
 CC 1 (ubiq1) intron sequence into BBTv promoter constructs enhances promoter
 CC activity. The present sequence represents the DNA sequence of S1 promoter
 CC fragment derived from BBTv-S1 and includes an open reading frame fragment
 CC and an intergenic region fragment.
 XX
 SQ Sequence 593 BP; 189 A; 103 C; 133 G; 168 T; 0 other;

Query Match 9.8%; Score 57.4; DB 20; Length 593;
 Best Local Similarity 54.5%; Pred. No. 2.8e-09;
 Matches 115; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
 QY 4 gagtatttaattgcccctgttagaattgtttaagaacagggcattcagttcggacaaa 63
 DB 167 gattatttaattgattgatttagagggcatttaagagtaggggtgtagagagtactaaa 226
 QY 64 tacgaacccttagttatcttgggttcgaccatgtgcattgtcgtactctgtatttcccaatgtc 123
 DB 227 tataacctaattagtttagttgaattgatttaatacatatcatgttaattgtcagtgctaatctc 286
 QY 124 ctgctgattatttgaataatcagcagggacagataaaaactgtggaatatatttaaagtatg 183
 DB 287 atgcagaattctgtaaaatctccgaagatagaataagattattattgtttaaagaagga 346
 QY 184 tgcattcaattacacaaataccgcgcgc 214
 DB 347 aattattatgcactatgacaatcgtacgc 377

RESULT 4
 AAV71834
 ID AAV71834 standard; DNA; 1091 BP.
 XX
 AC AAV71834;
 XX
 DT 10-FEB-1999 (first entry)
 XX
 DE BBTv DNA II clone (2) nucleotide sequence.
 XX
 KW Banana bunchy top virus; BBTv; PCR technique; plant; tissue; banana;
 KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
 XX
 OS Banana bunchy top virus.
 XX
 PN US5846705-A.
 XX
 PD 08-DEC-1998.
 XX
 PF 06-APR-1995; 95US-0418071.
 XX
 PR 06-APR-1995; 95US-0418071.
 XX
 PA (BIOT-) DEV CENT BIOTECHNOLOGY.
 XX
 PI Soong T, Wu R, You L;
 XX
 WPI; 1999-059037/05.
 XX
 PT Nucleic acids having banana bunchy top virus component sequences -
 PT used to design primers for use in polymerase chain reaction
 PT detection of the virus
 XX
 PS Disclosure; Fig 12A-C; 27pp; English.
 XX
 CC This represents the nucleotide sequence of a banana bunchy top virus
 CC (BBTV) DNA II (clone 2) and represents a subgenomic DNA. The invention
 CC provides nucleic acid sequences associated with BBTv that can be used in
 CC a PCR technique for detecting BBTv. The nucleic acid sequences (AAV71830
 CC to AAV71833) are used as the basis for the construction of PCR primers,

CC to detect BBTv infection. The PCR technique is used for detecting BBTv
 CC in plant tissues (preferably banana, especially Musa species). The
 CC virus, one of the most important banana species viruses, causes phloem
 CC damage and is transmitted by aphids. PCR detection gives accurate,
 CC reliable and specific determination of absence or presence of the
 CC virus.
 XX
 SQ Sequence 1091 BP; 343 A; 224 C; 248 G; 276 T; 0 other;

Query Match 9.0%; Score 52.2; DB 20; Length 1091;
 Best Local Similarity 58.0%; Pred. No. 2.5e-07;
 Matches 111; Conservative 0; Mismatches 78; Indels 3; Gaps 1;
 QY 1 tttagattatttaattgcccctgttagaattgtttaagaacagggcattcagttcggac 60
 DB 744 tcagattattcgtgaattatgctgttatagaacaaataaagaatagattttataataataca 803
 QY 61 aaatacgaacc---ccttagttatcttgggttcgaccatgtgcattgtcgtactgtattg 117
 DB 804 aaatacgaaccatgtgtgattagaagaatggacaaatgtccatgtaattgttatggca 863
 QY 118 aatgtccctgctgattatttgaataatcagcagggacagataaaaactgtggaatttaa 177
 DB 864 aatgtgtgctgattatttgaataatcagcagggacagataaaaactgtggaatttaa 177
 QY 178 agtatgtgtcattc 190
 DB 924 gaaaggaaacttc 936

RESULT 5
 AAA38947
 ID AAA38947 standard; DNA; 1091 BP.
 XX
 AC AAA38947;
 XX
 DT 23-AUG-2000 (first entry)
 XX
 DE Banana bunchy top virus related nucleotide sequence #2.
 XX
 KW Banana bunchy top virus; BBTv; detection; ds.
 XX
 OS Banana bunchy top virus.
 XX
 PN TW360710-A.
 XX
 PD 11-JUN-1999.
 XX
 PF 30-JUN-1994; 94TW-0106105.
 XX
 PR 30-JUN-1994; 94TW-0106105.
 XX
 PA (BIOT-) DEV CENT BIOTECHNOLOGY.
 XX
 PI Wu R, You L, Song T;
 XX
 WPI; 2000-316145/27.
 XX
 PT Two circular single-stranded DNAs associated with banana bunchy top
 PT virus and detection of the virus -
 XX
 PS Claim 2; Page 1; 7pp; Chinese.
 XX
 CC The present invention describes two circular single-stranded DNAs
 CC associated with banana bunchy top virus (BBTV) and the encoded proteins.
 CC Also described is a method of detecting the BBTv virus using the
 CC polymerase chain reaction (PCR). The present sequence represents a
 CC BBTv related nucleotide sequence from the present invention.
 XX
 SQ Sequence 1091 BP; 345 A; 224 C; 247 G; 275 T; 0 other;

Query Match	9.0%;	Score 52.2;	DB 21;	Length 1091;
Best Local Similarity	58.0%;	Pred. No. 2.5e-07;		
Matches 112;	Conservative	0;	Mismatches 78;	Indels 3;
Gaps 1;				
1	ttagagtatttaaatatgcctgttgtagaatgtgttaagaacagggcattcagttcggac	60		
744	tcagattatctgaattatggcgttatagacaacaattaaagaatagagttttaataaataca	803		
61	aaatacgaacc---cctaagtattcttgggttcgaccatgtgcattctactcgtatttgc	117		
804	aaatacgaacatgtcgattagaaaaagatggcacaataatgccatgttaattgtatggca	863		
118	aatgtcctgcctgattatttgaataatcagcagggacagacaataaactgtggaatattaa	177		
864	aatgtgttcctgattatttgaataatcagcagggacagacaataaactgtggaatattaa	923		
178	agtatgtgtcatc 190			
924	gaaaggaacttc 936			
RESULT 6				
AAV71833	AAV71833 standard; DNA; 1096 BP.			
XX	AAV71833;			
XX	10-FEB-1999 (first entry)			
XX	BBTV DNA II clone (2-17) nucleotide sequence.			
XX	Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;			
XX	Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.			
XX	Banana bunchy top virus.			
Key	Location/Qualifiers			
FT	TATA_signal	1..7		
FT	stem_loop	/*tag= a		
FT	CDS	/*tag= b		
FT	polyA_signal	/*tag= c		
FT	polyA_signal	/*tag= d		
FT	polyA_signal	/*tag= e		
FT	polyA_signal	/*tag= f		
FT	polyA_signal	/*tag= g		
FT	polyA_signal	/*tag= h		
FT	polyA_signal	/*tag= i		
FT	polyA_signal	/*tag= j		
FT	polyA_signal	/*tag= k		
FT	polyA_signal	/*tag= l		
FT	polyA_signal	/*tag= m		
FT	polyA_signal	/*tag= n		
FT	polyA_signal	/*tag= o		
FT	polyA_signal	/*tag= p		
FT	polyA_signal	/*tag= q		
FT	polyA_signal	/*tag= r		
FT	polyA_signal	/*tag= s		
FT	polyA_signal	/*tag= t		
FT	polyA_signal	/*tag= u		
FT	polyA_signal	/*tag= v		
FT	polyA_signal	/*tag= w		
FT	polyA_signal	/*tag= x		
FT	polyA_signal	/*tag= y		
FT	polyA_signal	/*tag= z		
FT	polyA_signal	/*tag= aa		
FT	polyA_signal	/*tag= ab		
FT	polyA_signal	/*tag= ac		
FT	polyA_signal	/*tag= ad		
FT	polyA_signal	/*tag= ae		
FT	polyA_signal	/*tag= af		
FT	polyA_signal	/*tag= ag		
FT	polyA_signal	/*tag= ah		
FT	polyA_signal	/*tag= ai		
FT	polyA_signal	/*tag= aj		
FT	polyA_signal	/*tag= ak		
FT	polyA_signal	/*tag= al		
FT	polyA_signal	/*tag= am		
FT	polyA_signal	/*tag= an		
FT	polyA_signal	/*tag= ao		
FT	polyA_signal	/*tag= ap		
FT	polyA_signal	/*tag= aq		
FT	polyA_signal	/*tag= ar		
FT	polyA_signal	/*tag= as		
FT	polyA_signal	/*tag= at		
FT	polyA_signal	/*tag= au		
FT	polyA_signal	/*tag= av		
FT	polyA_signal	/*tag= aw		
FT	polyA_signal	/*tag= ax		
FT	polyA_signal	/*tag= ay		
FT	polyA_signal	/*tag= az		
FT	polyA_signal	/*tag= ba		
FT	polyA_signal	/*tag= bb		
FT	polyA_signal	/*tag= bc		
FT	polyA_signal	/*tag= bd		
FT	polyA_signal	/*tag= be		
FT	polyA_signal	/*tag= bf		
FT	polyA_signal	/*tag= bg		
FT	polyA_signal	/*tag= bh		
FT	polyA_signal	/*tag= bi		
FT	polyA_signal	/*tag= bj		
FT	polyA_signal	/*tag= bk		
FT	polyA_signal	/*tag= bl		
FT	polyA_signal	/*tag= bm		
FT	polyA_signal	/*tag= bn		
FT	polyA_signal	/*tag= bo		
FT	polyA_signal	/*tag= bp		
FT	polyA_signal	/*tag= bq		
FT	polyA_signal	/*tag= br		
FT	polyA_signal	/*tag= bs		
FT	polyA_signal	/*tag= bt		
FT	polyA_signal	/*tag= bu		
FT	polyA_signal	/*tag= bv		
FT	polyA_signal	/*tag= bw		
FT	polyA_signal	/*tag= bx		
FT	polyA_signal	/*tag= by		
FT	polyA_signal	/*tag= bz		
FT	polyA_signal	/*tag= ca		
FT	polyA_signal	/*tag= cb		
FT	polyA_signal	/*tag= cc		
FT	polyA_signal	/*tag= cd		
FT				

Query Match 7.8%; Score 45.6; DB 17; Length 1017;
 Best Local Similarity 56.4%; Pred. No. 4.6e-05;
 Matches 106; Conservative 0; Mismatches 79; Indels 3; Gaps 1;

QY 1 ttagagattttaaattgacctgttagaattgttttaagaacagggcattcagttcgac 60
 DB 726 tcagagtagtgtaattggtgtaataagacaggttaagaataagggttaaggtaataact 785

QY 61 aaatacgaacc---ccttagttatcttggttcgacatgtgcattgtactctgtattggc 117
 DB 786 agtatgagccatgtgtaattgcggagatgaataatcctctgttcattgattgtttgca 845

QY 118 aatgctcgtcgtgatttttgaataacagcagggagacagataaaactgtggaattattaa 177
 DB 846 aatgactcaccagatttgggaaataagtggaagatagaataaaataaatttcgtgttga 905

QY 178 agtatgtg 185
 DB 906 aaactctg 913

RESULT 10
 AAS70474
 ID AAS70474 standard; cDNA; 2885 BP.
 XX
 AC AAS70474;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #6278.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG06287.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 1; SEQ ID No 6278; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2885 BP; 1144 A; 623 C; 512 G; 606 T; 0 other;

Query Match 6.3%; Score 36.8; DB 23; Length 2885;
 Best Local Similarity 56.7%; Pred. No. 0.092;
 Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 34 gtttaagaacagggcattcagttcgacacaaatcagaccccttagttattctgtggttcgac 93
 DB 1333 gttctggcagggaatcaggaagacaaagaaataaagggtattagttaggaaagag 1392

QY 94 catgtgcattgtactgtatttgcacatgtcctgcctgattttgaaatcagcaggggac 153
 DB 1393 gaagtcaattgtcctgtttgcagatgacatgttttatttagaataatccccatcgac 1452

RESULT 11
 AAS68521
 ID AAS68521 standard; cDNA; 1935 BP.
 XX
 AC AAS68521;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #4325.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG04334.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 1; SEQ ID No 4325; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC disorders involving aberrant protein expression or biological activity.

CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1935 BP; 394 A; 325 C; 414 G; 802 T; 0 other;

Query Match 6.0%; Score 35; DB 23; Length 1935;
 Best Local Similarity 56.5%; Pred. No. 0.31;
 Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 34 gtttaaacagggcattcagttcgacaaatacgaaccccttagttatcttggttcgac 93
 DB 525 GTTCTGCCAGGCAATCATCGGGCGGAGAAAGAAATTAAGGGTATTCATTTAGGAAGAG 466
 QY 94 catgtcgatctactgtatttgcacatgtcctgcctgatttggaaatcagca 148
 DB 465 GAAGTCAATTCCTCTGTTGCAGATGACATGCTATATTAGAAAACCCCA 411

RESULT 14
 AAV24091
 ID AAV24091 standard; DNA; 1103 BP.
 XX
 AC AAV24091;
 DT 11-AUG-1998 (first entry)
 DE Banana bunchy top virus component 1 DNA sequence.
 DE BBTV; probe; diagnostic primer; component 1; ss.
 KW Banana bunchy top virus.
 XX US5756708-A.
 XX 26-MAY-1998.
 XX 24-FEB-1994; 94US-0202186.
 XX 24-FEB-1994; 94US-0202186.
 XX (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
 XX Burns TM, Dale JL, Harding RM, Karan M;
 XX WPI; 1998-321636/28.
 XX Isolated Banana bunchy top virus DNA - useful as diagnostic probes
 XX and primers for producing virus-resistant plants
 XX Disclosure; Column 25-28; 59pp; English.
 XX This sequence represents a DNA fragment from the component 1 of
 CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
 CC diagnostic probes or primers or can be inserted into plants or other
 CC organisms e.g. to produce virus-resistant plants or to act as promoters,
 CC enhancers or termination signals.

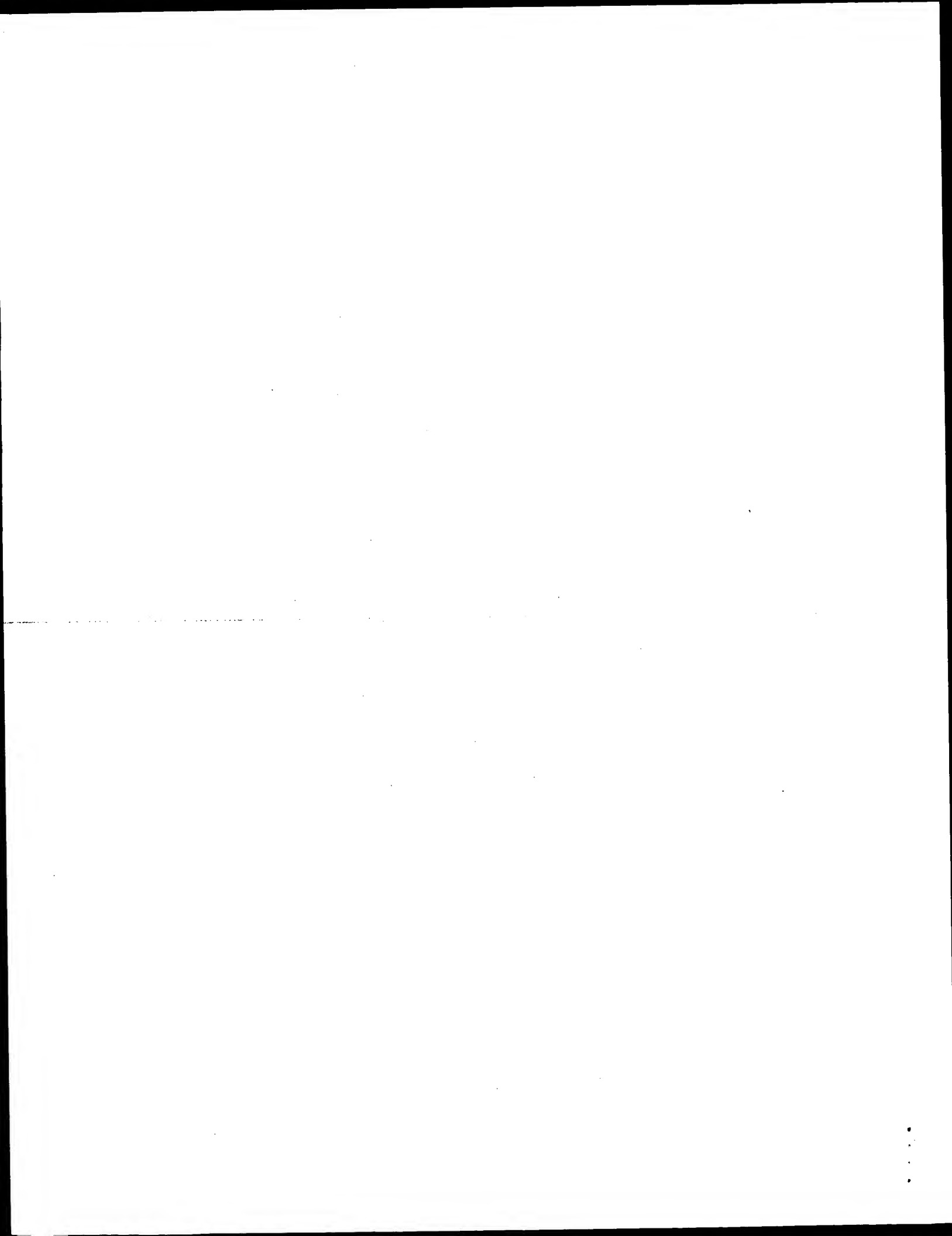
XX
 SQ Sequence 1103 BP; 355 A; 192 C; 277 G; 279 T; 0 other;
 Query Match 5.9%; Score 34.6; DB 19; Length 1103;
 Best Local Similarity 55.3%; Pred. No. 0.32;
 Matches 89; Conservative 0; Mismatches 69; Indels 3; Gaps 1;
 QY 4 gagtatttaattatgcctgttagaattgtttaagaacagggcattcagttcgacaaa 63
 DB 792 gaattatttaactatgcttattagaagaatttaaaatggaattattcaaaagcgggaaa 851
 QY 64 tagaacccttagttatcttgggttcgacacatgtgcattctgatttgcgaatgtc 123
 DB 852 tatgaaccctttgaaaattg---tagaatgtggaatcattgttaatggttaacttc 908
 QY 124 ctgctgattatttgaataatcagcaggagacagaataaaact 164
 DB 909 ctccgaaggaaggaattctttctggaagatcgataaagct 949

RESULT 15
 AAV24093
 ID AAV24093 standard; DNA; 1105 BP.
 XX
 AC AAV24093;
 DT 11-AUG-1998 (first entry)
 DE Banana bunchy top virus component 1 DNA sequence.
 DE BBTV; probe; diagnostic primer; component 1; ss.
 KW Banana bunchy top virus.
 XX OS
 XX US5756708-A.
 XX 26-MAY-1998.
 XX 24-FEB-1994; 94US-0202186.
 XX 24-FEB-1994; 94US-0202186.
 XX (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
 XX Burns TM, Dale JL, Harding RM, Karan M;
 XX WPI; 1998-321636/28.
 XX Isolated Banana bunchy top virus DNA - useful as diagnostic probes
 XX and primers for producing virus-resistant plants
 XX Disclosure; Column 29-30; 59pp; English.
 XX This sequence represents a DNA fragment from the component 1 of
 CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
 CC diagnostic probes or primers or can be inserted into plants or other
 CC organisms e.g. to produce virus-resistant plants or to act as promoters,
 CC enhancers or termination signals.
 XX
 SQ Sequence 1105 BP; 359 A; 197 C; 279 G; 270 T; 0 other;

Query Match 5.9%; Score 34.6; DB 19; Length 1105;
 Best Local Similarity 55.3%; Pred. No. 0.32;
 Matches 89; Conservative 0; Mismatches 69; Indels 3; Gaps 1;
 QY 4 gagtatttaattatgcctgttagaattgtttaagaacagggcattcagttcgacaaa 63
 DB 794 gaattatttaactatgcttattagaagaatttaaaatggaattattcaaaagcgggaaa 853
 QY 64 tagaacccttagttatcttgggttcgacacatgtgcattctgatttgcgaatgtc 123

Db 854 tatgaacccggttttgaaaattg---tagaatatgtggaagtcacatgtaatggctaacttc 910
QY 124 ctgacctgattattttgaaaatcacagcaggagacagaataaaaaact 164
Db 911 ctccgaagggaaggaatcttttcagaaagatccgaataaaagct 951

Search completed: June 19, 2002, 16:51:12
Job time: 17530 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:49:15 ; Search time 247.21 Seconds
(without alignments)
579.282 Million cell updates/sec

Title: US-09-462-955B-1_COPY_409_991

Perfect score: 583
Sequence: 1 ttagagtatttaattatgc.....cccccaacctgtctaacc 583

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	493	84.6	1291	4	US-09-462-975-1
2	58	9.9	1106	2	US-08-418-071-3
3	52.2	9.0	1091	2	US-08-418-071-5
4	52.2	9.0	1096	2	US-08-418-071-4
5	45.6	7.8	1017	4	US-08-793-634B-6
6	34.6	5.9	1103	1	US-08-202-186-16
7	34.6	5.9	1105	1	US-08-202-186-18
8	34	5.8	1110	1	US-08-202-186-14
9	34	5.8	1111	1	US-08-202-186-12
10	33	5.7	1104	1	US-08-202-186-17
11	32.4	5.6	982	3	US-08-973-068-28
12	32.4	5.6	1109	1	US-08-202-186-13
13	32.4	5.6	1111	1	US-08-202-186-9
14	32.4	5.6	1111	1	US-08-202-186-24
15	32.2	5.5	4411529	4	US-09-103-840A-1
C 16	31	5.3	31	4	US-09-462-975-2
C 17	30.8	5.3	1110	1	US-08-202-186-11
18	30.8	5.3	1111	1	US-08-202-186-10
19	30.8	5.3	1111	1	US-08-202-186-15
20	30	5.1	265	3	US-08-289-222E-5
21	30	5.1	265	4	US-09-054-526B-5
22	30	5.1	885	4	US-08-858-207A-6
C 23	30	5.1	2088	4	US-09-351-414-3
24	30	5.1	2272	1	US-08-482-577B-1
25	30	5.1	2272	3	US-08-289-222E-2
26	30	5.1	2272	4	US-09-218-176-1
27	30	5.1	2272	4	US-09-054-526B-2

C 28	29.4	5.0	859	3	US-09-280-409-2	Sequence 2, Appli
29	29.4	5.0	1022	4	US-08-793-634B-2	Sequence 2, Appli
30	29.4	5.0	1414	4	US-08-686-968C-228	Sequence 228, App
31	29.4	5.0	8920	2	US-08-446-855A-1	Sequence 1, Appli
32	29.4	5.0	8920	4	US-09-150-741-1	Sequence 1, Appli
33	29	5.0	9468	1	US-08-325-547-10	Sequence 10, Appli
C 34	28.8	4.9	10607	1	US-08-078-090-3	Sequence 3, Appli
35	28.6	4.9	2899	2	US-08-624-581-2	Sequence 2, Appli
36	28.6	4.9	2917	2	US-08-624-581-3	Sequence 3, Appli
37	28.6	4.9	2959	2	US-08-624-581-1	Sequence 1, Appli
38	28.4	4.9	7832	4	US-09-004-838-94	Sequence 94, Appli
C 39	28	4.8	2483	1	US-08-526-964-1	Sequence 1, Appli
C 40	28	4.8	2483	2	US-08-946-617-1	Sequence 5, Appli
C 41	28	4.8	2483	3	US-09-031-897-1	Sequence 1, Appli
C 42	28	4.8	5203	4	US-09-257-770-1	Sequence 25, Appli
43	28	4.8	19932	2	US-08-477-451-25	Sequence 25, Appli
44	27.8	4.8	1701	3	US-08-599-968-2	Sequence 2, Appli
C 45	27.8	4.8	2437	3	US-08-904-284-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-462-975-1
; Sequence 1, Application US/09462975
; Patent No. 6303345
; GENERAL INFORMATION:
; APPLICANT: Rohde, Wolfgang
; APPLICANT: Becker, Dieter
; APPLICANT: Randles, John W.
; APPLICANT: Hehn, Alain
; APPLICANT: Salmini, Francesco
; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER
; FILE REFERENCE: 23232.0003U1
; CURRENT APPLICATION NUMBER: US/09462,975
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/EP98/04345
; PRIOR FILING DATE: 1998-07-13
; PRIOR APPLICATION NUMBER: 19730502.4
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1291
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:/note=synthetic
; OTHER INFORMATION: construct
US-09-462-975-1

Query Match 84.6%; Score 493; DB 4; Length 1291;
Best Local Similarity 100.0%; Pred. No. 3.5e-165;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttagagtatttaattatgcctgttagaagtgttaagaacaggcatttcaggac 60
|||||
Db 799 ttagagtatttaattatgcctgttagaagtgttaagaacaggcatttcaggac 858
|||||

Qy 61 aaatcacaccccttagttatcttggttcacacatgtgcattgtatttgcacat 120
|||||
Db 859 aaatcacaccccttagttatcttggttcacacatgtgcattgtatttgcacat 918
|||||

Qy 121 gtctgctgattatttgaataatcacaggagacagataaaactgtggaattataagt 180
|||||
Db 919 gtctgctgattatttgaataatcacaggagacagataaaactgtggaattataagt 978
|||||

Qy 181 atgtgtcatctaaattac 240
|||||
Db 979 atgtgtcatctaaattac 1038
|||||

Qy 241 aatctctgccagccgaagcctgaggtgctaccgcgcgaagccggaacaata 300
 Db 1039 aatctctgccagccgaagcctgaggtgctaccgcgcgaagccggaacaata 1098
 Qy 301 tgaatcgagttatggcgccgcaacaataaaagattccatttggataaagcaaatctgt 360
 Db 1099 tgaatcgagttatggcgccgcaacaataaaagattccatttggataaagcaaatctgt 1158
 Qy 361 tactttgctgagtcgacgcgaacacatttccacgtccacaaatccaggtgagtagctgc 420
 Db 1159 tactttgctgagtcgacgcgaacacatttccacgtccacaaatccaggtgagtagctgc 1218
 Qy 421 tggagaagaagccgcaagcctctatctacgtccatttgttttgcgcgatcggaacgc 480
 Db 1219 tggagaagaagccgcaagcctctatctacgtccatttgttttgcgcgatcggaacgc 1278
 Qy 481 tgagttgatctgg 493
 Db 1279 tgagttgatctgg 1291

RESULT 2
 US-08-418-071-3
 ; Sequence 3, Application US/08418071
 ; Patent No. 5846705
 ; GENERAL INFORMATION:
 ; APPLICANT: Wu, Rey-Yuh
 ; APPLICANT: You, Li-Ru
 ; APPLICANT: Soong, Tai-Seng
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
 ; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECT
 ; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Neave
 ; STREET: 1251 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States
 ; ZIP: 10020
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 5.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/418,071
 ; FILING DATE: 06-APR-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Haley Jr., James F.
 ; REGISTRATION NUMBER: 27,794
 ; REFERENCE/DOCKET NUMBER: DCB-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 596-9000
 ; TELEFAX: (212) 596-9000
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1106 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: circular
 ; MOLECULE TYPE: genomic DNA
 ; DESCRIPTION: desc= "BBTV DNA I (clone 7-4-2)"
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)
 ; US-08-418-071-3

Query Match 9.9%; Score 58; DB 2; Length 1106;
 Best Local Similarity 57.9%; Pred. No. 1.4e-10;

Matches 103; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
 Qy 4 gagtattaaattatgcccctgttagaattgtttaagaacagggcatttcggaacaa 63
 Db 749 GATTATTAAATATGATCTAATACAGGCAATTAAAGGATAGGTTATAGAGAGTACTAAA 808
 Qy 64 tacgaaccccttagttatcttgggttcgacctggtgcatgtactctgttatttgccaatctc 123
 Db 809 TACAACACCATAAAGATAGTTGAATTAGTAAATACATGTAATCTCATGCGCAATTTC 868
 Qy 124 ctgctgattattgaaatcagcagggcaggaataaaactgtggaatatttaaagta 181
 Db 869 ATGCTGTACTTCTGTAAATCTCCGAAGATCGAATAAATAATCATTATTATGCTGAAGAA 926

RESULT 3
 US-08-418-071-5
 ; Sequence 5, Application US/08418071
 ; Patent No. 5846705
 ; GENERAL INFORMATION:
 ; APPLICANT: Wu, Rey-Yuh
 ; APPLICANT: You, Li-Ru
 ; APPLICANT: Soong, Tai-Seng
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
 ; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETE
 ; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Neave
 ; STREET: 1251 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States
 ; ZIP: 10020
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 5.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/418,071
 ; FILING DATE: 06-APR-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Haley Jr., James F.
 ; REGISTRATION NUMBER: 27,794
 ; REFERENCE/DOCKET NUMBER: DCB-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 596-9000
 ; TELEFAX: (212) 596-9090
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1091 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: circular
 ; MOLECULE TYPE: subgenomic DNA
 ; DESCRIPTION: /desc "BBTV DNA II (clone 2)"
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)
 ; US-08-418-071-5

Query Match 9.0%; Score 52.2; DB 2; Length 1091;
 Best Local Similarity 58.0%; Pred. No. 1.6e-08;
 Matches 112; Conservative 0; Mismatches 78; Indels 3; Gaps 1;

Qy 1 ttacagattttaaattatgcccctgttagaattgtttaagaacagggcatttcggtcgac 60
 Db 744 TCAGATTATCTGAATTATGCGCTTATAGACAAATTAAGATAGAGTTTAAATAATACA 803

QY 61 aaatcagaacc---ccttagttatcttgggttcgaccatgtgcactgactcgtattgcc 117
|||||
Db 804 AAATACGAACCATGCTGATTAGAAAAGATGACAAAATGTCATGTAATTGATGGCA 863
QY 118 aatgctcctgcctgattattgaaatcagcaggacagacataaaactgtggaatttaa 177
|||||
Db 864 AATGTCCTGCTGATTATTGTAATAATTCAGAAAGATAGATAAAAAATAATTAATTGTTGA 923
QY 178 agtatgtgtcatc 190
| | | | |
Db 924 GAAAGGAAACTTC 936

RESULT 4

US-08-418-071-4
; Sequence 4, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTI
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1096 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-418-071-4

Query Match 9.0%; Score 52.2; DB 2; Length 1096;
Best Local Similarity 58.0%; Pred. No. 1.6e-08;
Matches 112; Conservative 0; Mismatches 78; Indels 3; Gaps 1;

QY 1 tttagagtatttaaatatgcctgttagaatgtgttaagaacagggcattcagttcggac 60
| | | | |
Db 749 TCAGATTATCTGATTTATGCGCTTATAGACAAATTAAGAATAGAGTTTAAATAATACA 808
QY 61 aaatcagaacc---ccttagttatcttgggttcgaccatgtgcactgactcgtattgcc 117
|||||

Db 809 AAATACGAACCATGCTGATTAGAAAAGATGACAAAATGTCATGTAATTGATGGCA 868
QY 118 aatgctcctgcctgattattgaaatcagcaggacagacataaaactgtggaatttaa 177
|||||
Db 869 AATGTCCTGCTGATTATTGTAATAATTCAGAAAGATAGATAAAAAATAATTAATTGTTGA 928
QY 178 agtatgtgtcatc 190
| | | | |
Db 929 GAAAGGAAACTTC 941

RESULT 5

US-08-793-634B-6
; Sequence 6, Application US/08793634B
; Patent No. 6211431
; GENERAL INFORMATION:
; APPLICANT: Boevink, Petra C.
; APPLICANT: Surin, Brian P.
; APPLICANT: Keese, Paul K.
; APPLICANT: Chu, Paul W.G.
; APPLICANT: Waterhouse, Peter M.
; APPLICANT: Khan, Rafiqul I.
; APPLICANT: Larkin, Phillip J.
; APPLICANT: Taylor, William C.
; APPLICANT: Marchall, Jerry S.
; TITLE OF INVENTION: NOVEL PLANT PROMOTERS AND USES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,634B
; FILING DATE: June 9, 1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1017 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-793-634B-6

Query Match 7.8%; Score 45.6; DB 4; Length 1017;
Best Local Similarity 56.4%; Pred. No. 3.3e-06;
Matches 106; Conservative 0; Mismatches 79; Indels 3; Gaps 1;

QY 1 tttagagtatttaaatatgcctgttagaatgtgttaagaacagggcattcagttcggac 60
| | | | |
Db 726 TCAGAGTATGTGAATTAATGTTGTAATAGACAGGTTAAGAAATAGGTAATGGTAATGACT 785
QY 61 aaatcagaacc---ccttagttatcttgggttcgaccatgtgcactgactcgtattgcc 117
| | | | |
Db 786 AAGTATGAGCCATGCTGTAATGCGGGATGATAATCATCTCTGTCAATGTAATGTTGTTGA 845
QY 118 aatgtcctgcctgattattgaaatcagcaggacagacataaaactgtggaatttaa 177
| | | | |

Db 846 AATGACTCCAGATGGGAAAAATTAAGTCAAGATAGATAAAATTAATTCGTTGTTGA 905
Qy 178 agtatgtg 185
Db 906 AAACCTCTG 913

RESULT 6
US-08-202-186-16
; Sequence 16, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-16

Query Match 5.9%; Score 34.6; DB 1; Length 1103;
Best Local Similarity 55.3%; Pred. No. 0.028;
Matches 89; Conservative 0; Mismatches 69; Indels 3; Gaps 1;
Qy 4 gagtattaaattatgcctttagaattgtttaagaacaggcattcagttcggaacaa 63
Db 792 GAATATTTAAACTAGCTGTTTATTAAGAAGATTTAAAAATGGAATTATTCAAAAGCGGAAA 851
Qy 64 tacgaacccttagttatcttgggttcgaccatgtgcattgactgtatttgccaatgtc 123
Db 852 TATGAACCCGTTTGTGAAAATTG---TAGAATATGTGGAGTCATTGTAATGGCTAACTTC 908
Qy 124 ctgcctgattatttgaataatcagcaggacagataaaact 164
Db 909 CTTCGAGGAGGAGGAACTTTTCTGAGATCGAATAAAGCT 949

US-08-202-186-16
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
US-08-202-186-16

; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-18

Query Match 5.9%; Score 34.6; DB 1; Length 1105;
Best Local Similarity 55.3%; Pred. No. 0.028;
Matches 89; Conservative 0; Mismatches 69; Indels 3; Gaps 1;
Qy 4 gagtattaaattatgcctttagaattgtttaagaacaggcattcagttcggaacaa 63
Db 794 GAATATTTAAACTAGCTGTTTATTAAGAAGATTTAAAGATTTAATTCAAAAGCGGAAA 853
Qy 64 tacgaacccttagttatcttgggttcgaccatgtgcattgactgtatttgccaatgtc 123
Db 854 TATGAACCCGTTTGTGAAAATTG---TAGAATATGTGGAGTCATTGTAATGGCTAACTTC 910
Qy 124 ctgcctgattatttgaataatcagcaggacagataaaact 164
Db 911 CTTCGAGGAGGAGGAACTTTTTCAGAGATCGAATAAAGCT 951

RESULT 8
US-08-202-186-14
; Sequence 14, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
US-08-202-186-14

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIKE
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1110 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
US-08-202-186-14

Query Match 5.8%; Score 34; DB 1; Length 1110;
Best Local Similarity 54.9%; Pred. No. 0.047;
Matches 89; Conservative 0; Mismatches 70; Indels 3; Gaps 1;

QY 4 gagtatttaattgcccctgttagaatgtgttaagaacagggcattcagttcgacaaa 63
DB 791 GATTATTTAAATTTATGTTTATTAGAGGAATTTAAGAATGGAATAATTCAAAGCGGAAA 850
QY 64 tacaaaccccttagttatcttgggttcgaccatgtgcattcgtatttgcaatgctc 123
DB 851 TATGAACCCCGTTTCAAGATTG---TAGATATGTCGAAGTCATTGTAATGGCTAACTTC 907
QY 124 ctgcctgattattgaaatacagcaggagacagaataaaactg 165
DB 908 CTTCGGAAGGAAGGAATCTTTCTGAGATCGAATAAAGTTG 949

RESULT 9
US-08-202-186-12
Sequence 12, Application US/08202186
Patent No. 5756708
GENERAL INFORMATION:
APPLICANT: KARAN, Mirko
APPLICANT: BURNS, Thomas M.
APPLICANT: DALE, James L.
APPLICANT: HARDING, Robert M.
TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIKE
TELECOMMUNICATION INFORMATION:

TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1111 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
US-08-202-186-12

Query Match 5.8%; Score 34; DB 1; Length 1111;
Best Local Similarity 54.9%; Pred. No. 0.047;
Matches 89; Conservative 0; Mismatches 70; Indels 3; Gaps 1;

QY 4 gagtatttaattgcccctgttagaatgtgttaagaacagggcattcagttcgacaaa 63
DB 792 GATTATTTAAATTTATGTTTATTAGAGGAATTTAAGAATGGAATAATTCAAAGCGGAAA 851
QY 64 tacgaaccccttagttatcttgggttcgaccatgtgcattcgtatttgcaatgctc 123
DB 852 TATGAACCCCGTTTGAAGATAG---TAGAATATGTCGAAGTCATTGTAATGGCTAACTTC 908
QY 124 ctgcctgattattgaaatacagcaggagacagaataaaactg 165
DB 909 CTTCGGAAGGAAGGAATCTTTCTGAGATCGAATAAAGTTG 950

RESULT 10
US-08-202-186-17
Sequence 17, Application US/08202186
Patent No. 5756708
GENERAL INFORMATION:
APPLICANT: KARAN, Mirko
APPLICANT: BURNS, Thomas M.
APPLICANT: DALE, James L.
APPLICANT: HARDING, Robert M.
TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIKE
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1104 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
US-08-202-186-17

Query Match 5.7%; Score 33; DB 1; Length 1104;


```

: NAME: JEFFERY, Donald D.
:
: REGISTRATION NUMBER: 19,980
: REFERENCE/DOCKET NUMBER: 71611/102 FIVE
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202 672 5300
: TELEFAX: 202 672 5399
: TELEX: 904136
:

```

[illegible]

```

14
RESULT
US-08-202-186-24
; Sequence 24, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202.186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular

```

```

;
; NAME/KEY: CDS
; LOCATION: 129...989
; US-08-202-186-24

Query Match          5.6%; Score 32.4; DB 1; Length 1111;
Best Local Similarity 54.3%; Pred. No. 0.17; Indels 3; Gaps 1;
Matches 88; Conservative 0; Mismatches 71;

QY 4 gagtattaaattatgccttggttagaatgtgttaagacaacagggcatttcggaacaaa 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 819 GAUTATTAAATTATGGCTTATTAGAGAAATTTAAGAAATGGAATAATTCAAGCGGGAAA 878

QY 64 tagcaaaccccttgattatcttggttcgaccatgtcatgtactcgtatttgcgaatgct 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 879 TAGCAACCCGTTTTGAGAGATAG---TGAATATATGTCGAAGHCATTGTAATGGCTAACTTC 935

QY 124 ctgcgtgattatttggaaaaatcacgacgggcagagaataaaaactg 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 936 CTTCGGAAGGAAGCAATCTTTCTGAGATCGAATAAAGTTG 977

```

```

RESULT 15
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```

Query Match      5.5%; Score 32.2; DB 4; Length 4411529;
Best Local Similarity 57.4%; Pred. No. 22;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 410 gagttagttgctgagagaagcgaagcctctctatctaccgtccattgtttttggcg 469
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3577728 GAGCGGCACGGCGGAACAAACACCGCGAGCTGCGGTCTCGGTGCGATCGGTGTCAC 35777689

Qy 470 gatcgagcgctgagttgatctggcgccaaaaacctctgct 510
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35777668 GCACCTGTTTCCGAGCGCGGCGACGAGCGGAATCGCAATGCT 35777628

```

Search completed: June 19, 2002, 16:09:16
Job time: 15014 sec

Thu Jun 20 06:56:39 2002

us-09-462-955b-1_copy_409_991.rni

Page 8

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 14:04:20 ; Search time 7489.97 Seconds
(without alignments)
1050.568 Million cell updates/sec

Title: US-09-462-955B-1_COPY_409_991

Perfect score: 583

Sequence: 1 ttagagtatttaattatgc.....ccccaacctctgtaccccc 583

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	40	6.9	512	12	AQ121628	AQ121628 HS_3089_A
C 2	38	6.5	2275	9	AF034173	AF034173 AF034173
C 3	36.8	6.3	498	12	AQ615572	AQ615572 HS_5144_B
C 4	36.4	6.2	787	12	CNS010B7	AL098749 Drosophil
C 5	36	6.2	507	12	A2444169	A2444169 IM02339F20
C 6	36	6.2	929	12	BH137680	BH137680 ENTPC13TR
C 7	35.6	6.1	318	12	AQ089903	AQ089903 HS_3001_A
C 8	35.6	6.1	1101	12	CNS016H0	AL106734 Drosophil
C 9	35.4	6.1	460	9	A1942816	AI942816 fc67b05_x
C 10	35.4	6.0	848	12	CNS00Q0X	AL085215 Arabidops
C 11	35.2	6.0	524	12	AQ746046	AQ746046 HS_2277_A
C 12	35	6.0	524	12	BH175883	BH175883 004_P_17-
C 13	35	6.0	524	12	CNS07TRP	AL612823 T3 end of
C 14	34.6	5.9	487	12	AQ418525	AQ418525 RPCI-11-2
C 15	34.6	5.9	555	12	AQ417598	AQ417598 RPCI-11-2
C 16	34.6	5.9	665	9	BG633121	BG633121 BB633121
C 17	34.6	5.9	689	12	AG167308	AG167308 Pan trogl

C 18	34.4	5.9	594	12	AZ397924	AZ397924 LM0163C12
C 19	34.4	5.9	898	9	AL667528	AL667528 AL667528
C 20	34.2	5.9	345	12	AZ401897	AZ401897 LM0168N22
C 21	34	5.8	442	12	AQ228962	AQ228962 HS_2013_B
C 22	34	5.8	443	12	AQ010299	AQ010299 HS_2172_B
C 23	34	5.8	1101	12	CNS00DBS	AL067144 Drosophil
C 24	33.8	5.8	410	12	AQ807176	AQ807176 HS_3249_A
C 25	33.8	5.8	673	12	AG167164	AG167164 Pan trogl
C 26	33.8	5.8	749	12	AZ331069	AZ331069 LM0056P10
C 27	33.8	5.8	772	12	AQ488982	AQ488982 RPCI-11-2
C 28	33.8	5.8	873	12	CNS012PA	AL101848 Drosophil
C 29	33.8	5.8	1174	10	BE779656	BE779656 601464455
C 30	33.6	5.8	526	12	AQ008775	AQ008775 RPCI11-22
C 31	33.6	5.8	403	10	BG006561	BG006561 QV4-GN014
C 32	33.6	5.8	418	12	AQ632834	AQ632834 RPCI-11-4
C 33	33.6	5.8	434	12	AQ728714	AQ728714 HS_5462_B
C 34	33.6	5.8	605	12	AQ481591	AQ481591 RPCI-11-2
C 35	33.6	5.8	658	12	AQ077737	AQ077737 CIT-HSP-2
C 36	33.4	5.7	283	9	AI065300	AI065300 TENU2187
C 37	33.4	5.7	348	12	AQ348508	AQ348508 RPCI11-11
C 38	33.4	5.7	411	9	AI226907	AI226907 ujl1b01.y
C 39	33.4	5.7	460	12	AQ624335	AQ624335 HS_2103_B
C 40	33.4	5.7	517	12	AQ565229	AQ565229 HS_5385_B
C 41	33.4	5.7	532	12	AQ565203	AQ565203 HS_5355_B
C 42	33.4	5.7	637	12	AQ377093	AQ377093 RPCI11-14
C 43	33.4	5.7	751	12	AQ745275	AQ745275 HS_5503_A
C 44	33.4	5.7	764	12	AQ240040	AQ240040 CIT-HSP-2
C 45	33.4	5.7	785	11	AK002866	AK002866 Mus muscu

ALIGNMENTS

RESULT 1
AQ121628/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

AQ121628
HS_3089_AL_B03_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3089 Col-5 Row-C, DNA sequence.
AQ121628
GSS:
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 512)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3089 row: C column: 5
Class: BAC ends
High quality sequence stop: 512.
Location/Qualifiers
1. 512
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-3089 Col-5 Row-C"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"


```

KEYWORDS      GSS.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
REFERENCE      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS        Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
TITLE          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
JOURNAL        1 (bases 1 to 787)
               Genoscope.
COMMENT        Direct Submission
               Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
               BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
               - Web : www.genoscope.cns.fr)
               Determination of this BAC-end sequence was carried out as part of a
               collaboration with the European Drosophila Genome Project (EDGP) -
               http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
               library (Dros BAC) was made by Alain Billaud at CEPH (Centre
               d'Etude du Polymorphisme Humain) with funding provided by a MRC
               project grant. The DNA was prepared from embryos by Alain Bucheton
               and Genevieve Pavan. It has been constructed in the vector
               pBeloBAC11.

FEATURES       Location/Qualifiers
Source         1..787
               /organism="Drosophila melanogaster"
               /plasmid="pBeloBAC11"
               /db_xref="taxon:7227"
               /clone_lib="DrosBAC"
               /clone="BACN03M07"
               /note="end : 77"

BASE COUNT    52 a 114 c 73 g 19 t 529 others
ORIGIN

Query Match   6.2%; Score 36.4; DB 12; Length 787;
Best Local Similarity 3.0%; Pred. No. 2.2;
Matches 10; Conservative 109; Mismatches 215; Indels 0; Gaps 0;

QY 189 tctaaattacacaaataccgccgcgcacgcgtatgtttacattctatgaatactct 248
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 425 KNTMMNNKNCMMNNKMMNNKMMNNKMMNNKMMNNKMMNNKMMNNKMMNNKMMNNMM 484
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 249 gccagggcgaaggcctggagtgctaccgcgcgaaggccgggaacaaatgaatga 308
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 485 KNNMKMMKMMNTNMKNTNNNNNNNNNNKNNKNNNNNNNNNNKNNKMMKMMKMMNN 544
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 309 gttatggggccacaaataaagattccatttggaataagaacgaatctgttacttgc 368
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 545 NMNTNMNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNNN 604
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 369 ttgcagtcacgcgaaccactttccacgtccacgaatccaggtagtagcttgaggag 428
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 605 NNMKMMKTMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNMM 664
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 429 aaagccgaagcctctatctaccgtccattgttttttcgcgcgacgcgcgtgaattga 488
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 665 TNNNNNNKNTMMNNKMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 724
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 489 tctgggcgaacaaacctgtctgaagtcgcctgct 522
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 725 MNKTMKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNK 758
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 5
LOCUS      AZ444169
DEFINITION LM0239F20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
ACCESSION AZ444169
VERSION    AZ444169.1 GI:10592981
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE      1 (bases 1 to 507)
AUTHORS        Dunn,D., Aovagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
               Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
               M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
               and Wright,D., Weiss,R.
TITLE          Mouse whole genome scaffolding with paired end reads from 10kb
               plasmid inserts
JOURNAL        Unpublished (2000)
COMMENT        Contact: Robert B. Weiss
               University of Utah Genome Center
               University of Utah
               Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
               84112, USA
               Tel: 801 585 5606
               Fax: 801 585 7177
               Email: rdunn@genetics.utah.edu
               Insert Length: 10000 Std Error: 0.00
               Plate: 0239 row: F column: 20
               Seq primer: CGTTGTAAACGAGCGCCAGT
               Class: plasmid ends
               High quality sequence stop: 507.

FEATURES       Location/Qualifiers
Source         1..507
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="UUGC1M0239F20"
               /clone_lib="Mouse 10kb plasmid UUGC1M library"
               /sex="Male"
               /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
               /note="Vector: PWD42nv; Purified genomic DNA from M.
               musculus C57BL/6J (male) was obtained from the Jackson
               Laboratory Mouse DNA Resource
               (http://www.jax.org/resources/documents/dnares/). The DNA
               was hydrodynamically sheared by repeated passage through a
               0.005 inch orifice at constant velocity. The sheared DNA
               was blunt end-repaired with T4 DNA polymerase and T4
               polynucleotide kinase. Adaptor oligonucleotides were
               ligated to the blunt ends in high molar excess. The
               adaptor DNA was purified and size-selected for a 9.5 to
               10.5 kb range using preparative agarose gel
               electrophoresis. Vector DNA was prepared from a derivative
               of PWD42 (gil4732114|gb|AF129072.1), a copy-number
               inducible derivative of plasmid R1. The vector was ligated
               with adaptors complementary to the insert adaptors and
               purified. The sheared, adaptor mouse DNA was annealed to
               adaptor vector DNA, and transformed into
               chemically-competent E. coli XL10-Gold (Stratagene) cells
               and selected for ampicillin resistance."

BASE COUNT    205 a 74 c 92 g 136 t
ORIGIN

Query Match   6.2%; Score 36; DB 12; Length 507;
Best Local Similarity 58.3%; Pred. No. 2.4;
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 96 tgtgcagtactcgtatttgcacatgtcctgcctgattatttgaataatcgaggagacag 155
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 364 TATTTCATGAACAGAAATGGGAAAGCATTCTCTCTATTAATTAATGATCAGTAGCATAT 305
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 156 aataaaactgtggaatatttaaagatgtgtcatctaaataacacaa 203
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 304 GATATTGTTCTGTCGAGGTTGAATCATCTTTTATTTTAAATAACAACAA 257
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
LOCUS      BH137680
DEFINITION ENTPL137R Entamoeba histolytica Sheared DNA Entamoeba histolytica
               genomic, DNA sequence.
ACCESSION   BH137680
VERSION     BH137680.1 GI:15096741

```



```

Db      247 AAAACCCA 255

RESULT 15
AQ417598
LOCUS   RP41-11-203C8-TV RP41-11 Homo sapiens genomic clone RP41-11-203C8,
DEFINITION
DNA sequence.
ACCESSION
AQ417598
VERSION 1 GI:4475441
KEYWORDS
SOURCE  human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 555)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE   Use of BAC End Sequences from Library RP41-11 for Sequence-Ready
Map Building
JOURNAL
COMMENT  Unpublished (1997)
Other_GSSs: RP41-11-203C8.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RP41-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES             Location/Qualifiers
     source            1..555
                        /organism="Homo sapiens"
                        /db_xref="GDB:757623"
                        /db_xref="taxon:9606"
                        /clone="RP41-11-203C8"
                        /clone_lib="RP41-11"
                        /sex="Male"
                        /cell_type="Lymphocytes"
                        /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RP4111 Human Male BAC Library"
BASE COUNT          192 a 114 c 110 g 139 t
ORIGIN

Query Match          5.9%; Score 34.6; DB 12; Length 555;
Best Local Similarity 54.3%; Pred. No. 7.1;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy  20 ccctgttagaattgttaagaacaggcattcagttcggacaaatacgaaccccttagtt 79
    ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  150 CTCGTGTAGATTCCGGTCGTCAGGGCAATCAGGCAGGAGAAAGAAATAAGGGGTGTTTC 209
    ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  80 atcttggttcgacattgctgactgctgattgctgcaatgctcctgctgattattga 139
    ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  210 AATTAGGAAAGAGAGAGTCAATTTGTCCTGTTTGCAGATGACATGATTGTGATCTAC 269
    ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  140 aaatcagca 148
    ||| ||||| |||
Db  270 AAAACCCA 278

```

1

Result	Query	Score	Match	Length	DB	ID	Description
No.							

Query Match 76.4%; Score 291; DB 6; Length 1291;
Best Local Similarity 100.0%; Pred. No. 1.4e-75;

1

Query Match	9.1%; Score 34.8; DB 9; Length 40709;	
Best Local Similarity	52.8%; Pred. No. 18;	
Matches 75; Conservative 0; Mismatches 67; Indels 0; Gaps 0;		
QY 72	ctaccgcccgaaggcgggaacaataatgaatcgaatgtatggcgccgcccacaataaaag 131	
Db 31651	CTTCCAGTCCCGAGCTGFGAATAATAATGCCATTGTTGAAGCCAAAGAAAAAAA 31592	
QY 132	attcattcttgataaagaacgaatctgttactttgcttgcagtgcaacgcaacacatttcca 191	
Db 31591	ATTCCATTGTTCCCTGACTACTTTTATTTATTTATTTTGTGACAGAGATCTCACTC 31532	
QY 192	cgtcaccaatccagggtgagtag 213	
Db 31531	CGTCGCCAGGCTGGAGTGCAG 31510	
RESULT 7		
AC000026/c		
LOCUS	AC000026 126312 bp DNA linear PRI 01-JUN-2000	
DEFINITION	Homo sapiens Chromosome 22q12 BAC Clone 58b8 In Meningioma Deletion Region, complete sequence.	
ACCESSION	AC000026	
VERSION	AC000026.3 GI:5881547	
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1. (bases 1 to 126312)	
.TITLE	Pan,H., Hartman,K., Willingham,D. and Roe,B.A.	
JOURNAL	Homo sapiens Chromosome 22q12 BAC Clone 58b8 In Meningioma Deletion Region	
REFERENCE	Unpublished	
AUTHORS	2. (bases 1 to 126312)	
JOURNAL	Kim,U.-J. and Simon,M.	
REFERENCE	Unpublished	
AUTHORS	3. (bases 1 to 126312)	
JOURNAL	Peyrard,M. and Dumanski,J.P.	
REFERENCE	Unpublished	
AUTHORS	4. (bases 1 to 126312)	
JOURNAL	Roe,B.A.	
REFERENCE	Direct Submission	
AUTHORS	Submitted (21-OCT-1996) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA	
5. (bases 1 to 126312)		
6. (bases 1 to 126312)		
7. (bases 1 to 126312)		
8. (bases 1 to 126312)		
9. (bases 1 to 126312)		
10. (bases 1 to 126312)		
11. (bases 1 to 126312)		
12. (bases 1 to 126312)		
13. (bases 1 to 126312)		
14. (bases 1 to 126312)		
15. (bases 1 to 126312)		
16. (bases 1 to 126312)		
17. (bases 1 to 126312)		
18. (bases 1 to 126312)		
19. (bases 1 to 126312)		
20. (bases 1 to 126312)		
21. (bases 1 to 126312)		
22. (bases 1 to 126312)		
23. (bases 1 to 126312)		
24. (bases 1 to 126312)		
25. (bases 1 to 126312)		
26. (bases 1 to 126312)		
27. (bases 1 to 126312)		
28. (bases 1 to 126312)		
29. (bases 1 to 126312)		
30. (bases 1 to 126312)		
31. (bases 1 to 126312)		
32. (bases 1 to 126312)		
33. (bases 1 to 126312)		
34. (bases 1 to 126312)		
35. (bases 1 to 126312)		
36. (bases 1 to 126312)		
37. (bases 1 to 126312)		
38. (bases 1 to 126312)		
39. (bases 1 to 126312)		
40. (bases 1 to 126312)		
41. (bases 1 to 126312)		
42. (bases 1 to 126312)		
43. (bases 1 to 126312)		
44. (bases 1 to 126312)		
45. (bases 1 to 126312)		
46. (bases 1 to 126312)		
47. (bases 1 to 126312)		
48. (bases 1 to 126312)		
49. (bases 1 to 126312)		
50. (bases 1 to 126312)		
51. (bases 1 to 126312)		
52. (bases 1 to 126312)		
53. (bases 1 to 126312)		
54. (bases 1 to 126312)		
55. (bases 1 to 126312)		
56. (bases 1 to 126312)		
57. (bases 1 to 126312)		
58. (bases 1 to 126312)		
59. (bases 1 to 126312)		
60. (bases 1 to 126312)		
61. (bases 1 to 126312)		
62. (bases 1 to 126312)		
63. (bases 1 to 126312)		
64. (bases 1 to 126312)		
65. (bases 1 to 126312)		
66. (bases 1 to 126312)		
67. (bases 1 to 126312)		
68. (bases 1 to 126312)		
69. (bases 1 to 126312)		
70. (bases 1 to 126312)		
71. (bases 1 to 126312)		
72. (bases 1 to 126312)		
73. (bases 1 to 126312)		
74. (bases 1 to 126312)		
75. (bases 1 to 126312)		
76. (bases 1 to 126312)		
77. (bases 1 to 126312)		
78. (bases 1 to 126312)		
79. (bases 1 to 126312)		
80. (bases 1 to 126312)		
81. (bases 1 to 126312)		
82. (bases 1 to 126312)		
83. (bases 1 to 126312)		
84. (bases 1 to 126312)		
85. (bases 1 to 126312)		
86. (bases 1 to 126312)		
87. (bases 1 to 126312)		
88. (bases 1 to 126312)		
89. (bases 1 to 126312)		
90. (bases 1 to 126312)		
91. (bases 1 to 126312)		
92. (bases 1 to 126312)		
93. (bases 1 to 126312)		
94. (bases 1 to 126312)		
9		

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/chromosome="22q12"
/map="22q12"
/clone="58b8"
BASE COUNT 31509 a 32141 c 32072 g 30590 t
ORIGIN

Query Match 9.1%; Score 34.8; DB 9; Length 126312;
Best local Similarity 52.8%; Pred. No. 18;
Matches 75; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 72 ctaccggcggaagcggggaacaatgaatcgagttatggcgggcccaataaaag 131
|||||
Db 117258 CTTCCAGTCCCGACGACTGTGAGAAATAAATTCATTTGTAAGCCAAAGAAAAAAA 117199
|||||

QY 132 attccatttgataagaacgaactctgttacttctgctgcagtcgacgcaaccactttcca 191
|||||
Db 117198 ATTCATTTGTTCCCTGACACACTTTTATTTATTTATTTTGTGACACAGAGTCCTCACTC 117139
|||||

QY 192 cgtcaccaatccaggtgagtag 213
|||||
Db 117138 CGTCGCCAGGCTGGAGTGAC 117117

RESULT 8
AC002059/c 173029 bp DNA linear PRI 01-JUN-2000
LOCUS
DEFINITION
Homo sapiens Chromosome 22q12 PAC Clone p704f1059q13 In Meningioma
Deletion Region, complete sequence.
ACCESSION
AC002059
VERSION
AC002059.3 GI:4559399
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 173029)
Pan,H., Malaj,E., Guillebault,D., Nguyen,T., Peyrard,M.,
Dumanski,J.P. and Roe,B.A.
Homo sapiens Chromosome 22q12 PAC Clone p704f1059q13 In Meningioma
Deletion Region
Unpublished
2 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (14-MAY-1997) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (17-FEB-1998) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (18-FEB-1998) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (21-MAR-1998) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
6 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (14-OCT-1998) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
7 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (17-OCT-1998) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
8 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (30-OCT-1998) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
9 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (22-NOV-1998) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
10 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (07-JAN-1999) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
11 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (14-JAN-1999) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
12 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (20-MAR-1999) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
13 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (23-MAR-1999) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
14 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (04-APR-1999) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
15 (bases 1 to 173029)
Roe,B.A.
Direct Submission
Submitted (22-APR-1999) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
16 (bases 1 to 173029)
Pan,H., Malaj,E., Guillebault,D., Nguyen,T., Peyrard,M.,
Dumanski,J.P. and Roe,B.A.
Direct Submission
Submitted (01-JUN-2000) Department Of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
Because these overlapping clones came from different libraries
there are numerous instances of insertions, deletions, and single
nucleotide polymorphisms in the overlapping regions below
AC000026(58b8) 8150 126312 (0) overlaps AC002059(p704f1059q13) 1
118230 (54799) AC002059(p704f1059q13) 77526 118235 (54794) overlaps
148038 (e81f2) 1 40709 (0) AC002059(p704f1059q13) 136102 173029 (0)
overlaps AC000041(p42hl) 1 36931 (8506).
FEATURES
Location/Qualifiers
1. .173029

```


Consensus quality: 159096 bases at least Q30
 Consensus quality: 165302 bases at least Q20
 Estimated insert size: 150759; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation

```

* * NOTE: Estimated insert size may differ from sequence length
* * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* * NOTE: This is a 'working draft' sequence. It currently
* * consists of 61 contigs, the true order of the pieces
* * is not known and their order in this sequence record is
* * arbitrary. Gaps between the contigs are represented as
* * runs of N, but the exact sizes of the gaps are unknown.
* * This record will be updated with the finished sequence
* * as soon as it is available and the accession number will
* * be preserved.

```

*	1	14312:	contig of 14312 bp in length
*	14313	14412:	gap of unknown length
*	14413	21907:	contig of 7495 bp in length
*	21908	22007:	gap of unknown length
*	22008	29349:	contig of 7342 bp in length
*	29350	29449:	gap of unknown length
*	29450	33945:	contig of 4496 bp in length
*	33946	34045:	gap of unknown length
*	34046	38558:	contig of 4513 bp in length
*	38559	38658:	gap of unknown length
*	38659	42850:	contig of 4192 bp in length
*	42851	42950:	gap of unknown length
*	42951	48562:	contig of 5612 bp in length
*	48563	48662:	gap of unknown length
*	48663	53475:	contig of 4813 bp in length
*	53475	53575:	gap of unknown length
*	53576	58890:	contig of 5315 bp in length
*	58891	58990:	gap of unknown length
*	58991	63577:	contig of 4587 bp in length
*	63578	63677:	gap of unknown length
*	63678	68788:	contig of 5111 bp in length
*	68789	68888:	gap of unknown length
*	68889	72330:	contig of 3742 bp in length
*	72631	72730:	gap of unknown length
*	72731	76336:	contig of 3606 bp in length
*	76337	76436:	gap of unknown length
*	76437	79855:	contig of 3419 bp in length
*	79856	79955:	gap of unknown length
*	79956	84438:	contig of 4483 bp in length
*	84439	84538:	gap of unknown length
*	84539	87215:	contig of 2677 bp in length
*	87216	87315:	gap of unknown length
*	87316	90687:	contig of 3372 bp in length
*	90688	90787:	gap of unknown length
*	90788	94269:	contig of 3482 bp in length
*	94270	94369:	gap of unknown length
*	94370	96285:	contig of 1916 bp in length
*	96286	96385:	gap of unknown length
*	96386	100073:	contig of 3688 bp in length
*	100074	100173:	gap of unknown length
*	100174	102829:	contig of 2656 bp in length
*	102830	102929:	gap of unknown length
*	102930	105029:	contig of 2100 bp in length
*	105030	105129:	gap of unknown length
*	105130	109381:	contig of 4252 bp in length
*	109382	109481:	gap of unknown length
*	109482	111493:	contig of 2012 bp in length
*	111494	111593:	gap of unknown length
*	111594	113967:	contig of 2374 bp in length
*	113968	114067:	gap of unknown length
*	114068	116094:	contig of 2027 bp in length
*	116095	116194:	gap of unknown length
*	116195	118535:	contig of 2341 bp in length
*	118536	118635:	gap of unknown length
*	118636	120330:	contig of 1695 bp in length
*	120331	120430:	gap of unknown length
*	120431	122526:	contig of 2096 bp in length

* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs, the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

* 11966 12065: gap of 100 bp
* 12066 13754: contig of 1689 bp in length
* 13755 13854: gap of 100 bp
* 13855 15223: contig of 1369 bp in length
* 15224 15323: gap of 100 bp
* 15324 16734: contig of 1411 bp in length
* 16735 16834: gap of 100 bp
* 16835 18271: contig of 1437 bp in length
* 18272 18371: gap of 100 bp
* 18372 19546: contig of 1175 bp in length
* 19547 19646: gap of 100 bp
* 20911 20910: contig of 1264 bp in length
* 20911 21010: gap of 100 bp
* 21011 22952: contig of 1942 bp in length
* 22953 23052: gap of 100 bp
* 23053 24765: contig of 1713 bp in length
* 24766 24865: gap of 100 bp
* 24866 26182: contig of 1317 bp in length
* 26183 26282: gap of 100 bp
* 26283 27734: contig of 1452 bp in length
* 27735 27834: gap of 100 bp
* 27835 29209: contig of 1375 bp in length
* 29210 29309: gap of 100 bp
* 29310 31590: contig of 2281 bp in length
* 31591 31690: gap of 100 bp
* 31691 33673: contig of 1983 bp in length
* 33674 33773: gap of 100 bp
* 33774 35628: contig of 1855 bp in length
* 35629 35728: gap of 100 bp
* 35729 37866: contig of 2138 bp in length
* 37867 37966: gap of 100 bp
* 37967 40549: contig of 2583 bp in length
* 40550 40649: gap of 100 bp
* 40650 42573: contig of 1926 bp in length
* 42576 42675: gap of 100 bp
* 42676 45614: contig of 2939 bp in length
* 45615 45714: gap of 100 bp
* 45715 48125: contig of 2411 bp in length
* 48126 48225: gap of 100 bp
* 48226 50633: contig of 2408 bp in length
* 50634 50733: gap of 100 bp
* 50734 53919: contig of 3186 bp in length
* 53920 54019: gap of 100 bp
* 54020 56516: contig of 2497 bp in length
* 56517 56616: gap of 100 bp
* 56617 59036: contig of 2420 bp in length
* 59037 59136: gap of 100 bp
* 59137 62691: contig of 3555 bp in length
* 62692 62791: gap of 100 bp
* 62792 65860: contig of 3069 bp in length
* 65861 65960: gap of 100 bp
* 65961 68910: contig of 2950 bp in length
* 68911 69010: gap of 100 bp
* 69011 72317: contig of 3307 bp in length
* 72318 72417: gap of 100 bp
* 72418 75535: contig of 3118 bp in length
* 75536 75635: gap of 100 bp
* 75636 78272: contig of 2637 bp in length
* 78273 78372: gap of 100 bp
* 78373 82077: contig of 3705 bp in length
* 82078 82177: gap of 100 bp
* 82178 86412: contig of 4235 bp in length
* 86413 86512: gap of 100 bp
* 86513 91421: contig of 4909 bp in length
* 91422 91521: gap of 100 bp
* 91522 96898: contig of 5377 bp in length
* 96899 96998: gap of 100 bp
* 96999 102230: contig of 5232 bp in length
* 102231 102330: gap of 100 bp
* 102331 106721: contig of 4391 bp in length
* 106722 106821: gap of 100 bp
* 106822 111047: contig of 4226 bp in length
* 111048 111147: gap of 100 bp

```

```

* 111148 116783: contig of 5636 bp in length
* 116784 116883: gap of 100 bp
* 116884 121808: contig of 4925 bp in length
* 121809 121908: gap of 100 bp
* 121909 127934: contig of 6026 bp in length
* 127935 128034: gap of 100 bp
* 128035 134776: contig of 6742 bp in length
* 134777 134876: gap of 100 bp
* 134877 142048: contig of 7172 bp in length
* 142049 142148: gap of 100 bp
* 142149 148978: contig of 6730 bp in length
* 148979 148978: gap of 100 bp
* 148979 157074: contig of 8096 bp in length
* 157075 157174: gap of 100 bp
* 157175 165913: contig of 8739 bp in length
* 165914 166013: gap of 100 bp
* 166014 177085: contig of 11072 bp in length
* 177086 177185: gap of 100 bp
* 177186 189278: contig of 12093 bp in length.

```

FEATURES
source

```

Location/Qualifiers
1..189278
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone_lib="RPC1-11 Human Male BAC"
1..1071
/note="assembly_fragment"
misc_feature
1172..2443
/note="assembly_fragment"
misc_feature
2544..4110
/note="assembly_fragment"
misc_feature
4211..5676
/note="assembly_fragment"
misc_feature
5777..6954
/note="assembly_fragment"
misc_feature
7055..8467
/note="assembly_fragment"
misc_feature
8568..10046
/note="assembly_fragment"
misc_feature
10147..11965
/note="assembly_fragment"
misc_feature
12066..13754
/note="assembly_fragment"
misc_feature
13855..15223
/note="assembly_fragment"
misc_feature
15324..16734
/note="assembly_fragment"
misc_feature
16835..18271
/note="assembly_fragment"
misc_feature
18372..19546

```

Query Match 8.8%; Score 33.6; DB 2; Length 189278;
Best Local Similarity 55.0%; Pred. No. 41;
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

```

QY 37 tgaatattccctgcccagccgagcctggtgaggtgctaccgccgagccgagccgagcaaa 96
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6594 TGCAGCACCTCCAGAGTCCAAATGACTGACGAGAGCTCCCTGCCAAAGAGCAGGAATAT 6653
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 97 tatgaatcgaggtatggcgccgcccacataaaagattccatttggtataagaacgaatct 156
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6654 TCATATTGGACTATTGGATCAGTCGACAGTATATACTTTTATATATTTTAAACACCTGAATT 6713
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 13
IAU86144
LOCUS
DEFINITION
Influenza A virus (A/Swine/Quebec/192/81 (SwQc81)) neuraminidase
mRNA, complete cds.
ACCESSION
U86144
VERSION
U86144.1 GI:4099318

1438 bp mRNA linear VRL 05-JAN-1999

* 9953 10603: contig of 651 bp in length
* 10604 10703: gap of 100 bp
* 10704 11455: contig of 752 bp in length
* 11456 11555: gap of 100 bp
* 11556 12282: contig of 727 bp in length
* 12283 12382: gap of 100 bp
* 12383 13126: contig of 744 bp in length
* 13127 13226: gap of 100 bp
* 13227 13922: contig of 696 bp in length
* 13923 14022: gap of 100 bp
* 14023 14755: contig of 733 bp in length
* 14756 14855: gap of 100 bp
* 14856 15599: contig of 744 bp in length
* 15600 15699: gap of 100 bp
* 15700 16435: contig of 736 bp in length
* 16436 16535: gap of 100 bp
* 16536 17265: contig of 730 bp in length
* 17266 17365: gap of 100 bp
* 17366 18082: contig of 717 bp in length
* 18083 18182: gap of 100 bp
* 18183 18909: contig of 727 bp in length
* 18910 19009: gap of 100 bp
* 19010 19713: contig of 710 bp in length
* 19720 19819: gap of 100 bp
* 19820 20556: contig of 737 bp in length
* 20557 20656: gap of 100 bp
* 20657 21396: contig of 740 bp in length
* 21397 21496: gap of 100 bp
* 21497 22238: contig of 742 bp in length
* 22239 22338: gap of 100 bp
* 22339 23068: contig of 730 bp in length
* 23069 23168: gap of 100 bp
* 23169 23894: contig of 726 bp in length
* 23895 23994: gap of 100 bp
* 23995 24733: contig of 739 bp in length
* 24734 24833: gap of 100 bp
* 24834 25567: contig of 734 bp in length
* 25568 25667: gap of 100 bp
* 25668 26394: contig of 727 bp in length
* 26395 26494: gap of 100 bp
* 26495 27239: contig of 745 bp in length
* 27240 27339: gap of 100 bp
* 27340 28073: contig of 734 bp in length
* 28074 28173: gap of 100 bp
* 28174 28923: contig of 750 bp in length
* 28924 29023: gap of 100 bp
* 29024 29753: contig of 730 bp in length
* 29754 29853: gap of 100 bp
* 29854 30594: contig of 741 bp in length
* 30595 30694: gap of 100 bp
* 30695 31431: contig of 737 bp in length
* 31432 31531: gap of 100 bp
* 31532 32252: contig of 721 bp in length
* 32253 32352: gap of 100 bp
* 32353 33097: contig of 745 bp in length
* 33098 33197: gap of 100 bp
* 33198 33949: contig of 752 bp in length
* 33950 34049: gap of 100 bp
* 34050 34781: contig of 732 bp in length
* 34782 34881: gap of 100 bp
* 34882 35598: contig of 717 bp in length
* 35599 35698: gap of 100 bp
* 35699 36438: contig of 740 bp in length
* 36439 36538: gap of 100 bp
* 36539 37289: contig of 751 bp in length
* 37290 37389: gap of 100 bp
* 37390 38086: contig of 697 bp in length
* 38087 38186: gap of 100 bp
* 38187 38916: contig of 730 bp in length
* 38917 39016: gap of 100 bp
* 39017 39763: contig of 747 bp in length
* 39764 39863: gap of 100 bp
* 39864 40616: contig of 753 bp in length

* 40617 40716: gap of 100 bp
* 40717 41450: contig of 734 bp in length
* 41451 41550: gap of 100 bp
* 41551 42274: contig of 724 bp in length
* 42275 42374: gap of 100 bp
* 42375 43106: contig of 732 bp in length
* 43107 43206: gap of 100 bp
* 43207 43960: contig of 754 bp in length
* 43961 44060: gap of 100 bp
* 44061 44814: contig of 754 bp in length
* 44815 44914: gap of 100 bp
* 44915 45641: contig of 727 bp in length
* 45642 45741: gap of 100 bp
* 45742 46462: contig of 721 bp in length
* 46463 46562: gap of 100 bp
* 46563 47306: contig of 744 bp in length
* 47307 47406: gap of 100 bp
* 47407 48135: contig of 729 bp in length
* 48136 48235: gap of 100 bp
* 48236 48972: contig of 737 bp in length
* 48973 49072: gap of 100 bp
* 49073 49807: contig of 735 bp in length
* 49808 49907: gap of 100 bp
* 49908 50639: contig of 732 bp in length
* 50640 50739: gap of 100 bp
* 50740 51474: contig of 735 bp in length
* 51475 51574: gap of 100 bp
* 51575 52306: contig of 732 bp in length
* 52307 52406: gap of 100 bp
* 52407 53137: contig of 731 bp in length
* 53138 53237: gap of 100 bp
* 53238 53953: contig of 716 bp in length
* 53954 54053: gap of 100 bp
* 54054 54775: contig of 722 bp in length
* 54776 54875: gap of 100 bp
* 54876 55593: contig of 718 bp in length
* 55594 55693: gap of 100 bp
* 55694 56418: contig of 725 bp in length
* 56419 56518: gap of 100 bp
* 56519 57267: contig of 749 bp in length
* 57268 57367: gap of 100 bp

Query Match 8.7%; Score 33; DB 2; Length 74854;

Best Local Similarity 49.7%; Pred. No. 61;

Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 29 acatcttatgaatctgcccagggccgaagcctggggagtgctaccgcgcgaagcc 88

Db 57085 AATACCTATTTATTTCTCCAGATAGCCAAAGGCTGTGAAGGCCACGAGTTCCACAGAGA 57026

QY 89 gggacaataatgaatcagtgatggcgcccaataaaagattccattggataaga 148

Db 57025 GAGAAAAGGATGATGCTGCTATTTGCCAGGTGCTCAGCAAGTTTCACGCAAAATCTGGGAA 56966

QY 149 acgaatctgtactgtctgctgagtcagcagcaaccactttccacgtcac 197

Db 56965 TCACCATAGTGTGATGATCTCTTTCCACGGATCTAATTACCACTTCAC 56917

RESULT 15

AC025672

LOCUS Homo sapiens chromosome 18 clone RP11-715F3 map 18, WORKING DRAFT
DEFINITION SEQUENCE, 20 unordered pieces.

ACCESSION AC025672

VERSION AC025672.2 GI:10198540

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 167178)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 18, clone RP11-715F3
Unpublished
2 (bases 1 to 167178)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,G., Liu,C., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Miengra,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced gi:7230282.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8225
Center clone name: 715_F3
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 159523 bases at least Q40
Consensus quality: 163250 bases at least Q30
Consensus quality: 164626 bases at least Q20
Insert size: 176000; agarose-fp
Quality coverage: 165278; sum-of-contigs
Quality coverage: 4.4 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 27917: contig of 27917 bp in length
* 27918 28017: gap of 100 bp
* 28018 29021: contig of 1004 bp in length
* 29022 29121: gap of 100 bp
* 29122 30937: contig of 1816 bp in length
* 30938 31037: gap of 100 bp
* 31038 33054: contig of 2017 bp in length
* 33055 33154: gap of 100 bp
* 33155 34403: contig of 1249 bp in length
* 34404 34503: gap of 100 bp
* 34504 36545: contig of 2042 bp in length
* 36546 36645: gap of 100 bp
* 36646 39583: contig of 2938 bp in length
* 39584 39683: gap of 100 bp

TITLE
JOURNAL
COMMENT

39684 44557: contig of 4874 bp in length
* 44558 44657: gap of 100 bp
* 44658 48993: contig of 4336 bp in length
* 48994 49093: gap of 100 bp
* 49094 53867: contig of 4774 bp in length
* 53868 53967: gap of 100 bp
* 53968 58665: contig of 4898 bp in length
* 58666 58765: gap of 100 bp
* 58766 65551: contig of 6786 bp in length
* 65552 65651: gap of 100 bp
* 65652 72916: contig of 7265 bp in length
* 72917 73016: gap of 100 bp
* 73017 83269: contig of 10253 bp in length
* 83270 83369: gap of 100 bp
* 83370 93557: contig of 10188 bp in length
* 93558 93657: gap of 100 bp
* 93658 103579: contig of 9922 bp in length
* 103580 103679: gap of 100 bp
* 103680 117213: contig of 13534 bp in length
* 117214 117313: gap of 100 bp
* 117314 133144: contig of 15831 bp in length
* 133145 133244: gap of 100 bp
* 133245 160275: contig of 27031 bp in length
* 160276 160375: gap of 100 bp
* 160376 167178: contig of 6803 bp in length.
FEATURES
source
1..167178
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-715F3"
/clone_lib="RPC1-11 Human Male BAC"
1..27917
/note="assembly_fragment"
clone_end:SP6
vector_side:left
28018..29021
/note="assembly_fragment"
29122..30937
/note="assembly_fragment"
31038..33054
/note="assembly_fragment"
33155..34403
/note="assembly_fragment"
34504..36545
/note="assembly_fragment"
36646..39583
/note="assembly_fragment"
39684..44557
/note="assembly_fragment"
44658..48993
/note="assembly_fragment"
49094..53867
/note="assembly_fragment"
53968..58665
/note="assembly_fragment"
58766..65551
/note="assembly_fragment"
65652..72916
/note="assembly_fragment"
73017..83269
/note="assembly_fragment"
83370..93557
/note="assembly_fragment"
93658..103579
/note="assembly_fragment"
103680..117213
/note="assembly_fragment"
117314..133144
/note="assembly_fragment"
133245..160275
/note="assembly_fragment"

misc_feature 160376..167178
 /note="assembly_fragment
 clone_end:T7
 vector_side:right"
 BASE COUNT 52172 a 31878 c 31581 g 49646 t 1901 others
 ORIGIN

Query Match 8.7%; Score 33; DB 2; Length 167178;
 Best Local Similarity 54.5%; Pred. No. 61;
 Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
 QY 99 tgaatcgagttatggcgggccacacataaaagattccatttggataagaacgaatctgt 158
 DB 15726 TGAAGAAGAGCTAGGGCTTCTCTGATTTAGACTTTGGATTAGGGAATGTTATGGCTGG 15785
 QY 159 tactttgttcagtgcaacgcaactttccacgtcccaatccagggtgagtagcttgc 218
 DB 15786 TTTGATCTATCGAGACCACACAAACGTTCTCCATGTCAGCAATAAAGCTGTTTGTTC 15845
 QY 219 t 219
 DB 15846 T 15846

Search completed: June 19, 2002, 15:48:13
 Job time: 13751 sec

Downloaded from <http://ajph.org/> on November 10, 2015

Human cDNA sequences
S. epidermidis genes
S. epidermidis genes
S. epidermidis genes
S. epidermidis genes
Human DJ889M15.3 h
Human protein encoding
Human polynucleotide
Human polynucleotide

DR WPI; 2001-318749/34.

RESULT 1
AAH16471

ID AAH16471 standard; cDNA: 2827 bp.

AA
AC
AAH16471;AA
DT 26-JUN-2001 (first entry)

XX
DE Human cDNA sequence SEQ ID NO:15487.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX
OS Homo sapiens.

XX
PN
EP1074617-A2XX
DD 07-FEB-2003

XX

FF 28-JUL-2000; 2000EP-0116126.
XX

PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253

PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000TR-0182757

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX

DR WPI; 2001-318749/34.


```

XX Kimmerly WJ;
XX WPI; 2001-316495/33.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX Claim 8; Page 1285-1287; 2188pp; English.
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
XX AAH55098 represent oligonucleotide sequences and primers which are used
XX in the exemplification of the present invention.
XX N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
XX Sequence 3295 BP; 971 A; 667 C; 456 G; 1201 T; 0 other;

Query Match      8.5%; Score 32.4; DB 22; Length 3295;
Best Local Similarity 53.1%; Pred. No. 0.88;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 200 atccaggtgagtagcttgctggagaagaagcgcgaagcctctatccgcgtccattgt 259
   || |||| || || || || || || || || || || || || || || || || || ||
DB 776 attaaagtattctcttagtatagaaaatgcctgagacaatgattgcccgacatttta 835

QY 260 ttctgcgcgattcgacggctgagttgatctggcgcgaacaaacctctctgaagtcogtgc 319
   || || || || || || || || || || || || || || || || || || || ||
DB 836 ttattggcaattgacacacggattggaactggcgtaattctcaattttcaactatagtc 895

QY 320 taagtataaa 329
   || || || ||
DB 896 tcattaccaa 905

RESULT 4
AAH54750
ID AAH54750 standard; DNA; 3518 BP.
XX
XX AAH54750;
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:4114.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX KW vaccination; endocarditis; ds.
XX
XX Staphylococcus epidermidis.
XX OS
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX

XX PA (GLAXO ) GLAXO GROUP LTD.
XX PI Kimmerly WJ;
XX
XX WPI; 2001-316495/33.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX Claim 8; Page 1808-1809; 2188pp; English.
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
XX AAH55098 represent oligonucleotide sequences and primers which are used
XX in the exemplification of the present invention.
XX N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
XX Sequence 3518 BP; 1091 A; 677 C; 495 G; 1255 T; 0 other;

Query Match      8.5%; Score 32.4; DB 22; Length 3518;
Best Local Similarity 53.1%; Pred. No. 0.91;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 200 atccaggtgagtagcttgctggagaagaagcgcgaagcctctatccgcgtccattgt 259
   || |||| || || || || || || || || || || || || || || || || || ||
DB 2590 attaaagtattctcttagtatagaaaatgcctgagacaatgattgcccgacatttta 2649

QY 260 ttctgcgcgattcgacggctgagttgatctggcgcgaacaaacctctgtaagtcogtgc 319
   || || || || || || || || || || || || || || || || || || || ||
DB 2650 ttattggcaattgacacacggattggaactggcgtaattctcaattttcaactatagtc 2709

QY 320 taagtataaa 329
   || || || ||
DB 2710 tcattaccaa 2719

RESULT 5
AAH54492
ID AAH54492 standard; DNA; 3985 BP.
XX
XX AAH54492;
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:3856.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX KW vaccination; endocarditis; ds.
XX
XX Staphylococcus epidermidis.
XX OS
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX

```


AAH99166 to AAH99904 encode the human proteins given in AAH25225 to AAH25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: antinflammatory; antineumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant; central nervous system; viricide; anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; antiagregant; haemostatic; vulnerary; antiulcer; osteopathic; dermatological; anti allergic; antiasthmatic; antididiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis

Claim 1; SEQ ID NO 4843; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

Sequence 5653 BP; 1262 A; 1680 C; 1632 G; 1079 T; 0 other;

Query Match 8.5%; Score 32.4; DB 22; Length 5653;
 Best Local Similarity 55.3%; Pred. No. 1.1;
 Matches 63; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
 QY 4 cccgcccgcacgcgtatgctttacatcttataatctcctccagggccgagcct 63
 DB 770 CCCAGGCGCTCCCATGCTCTTCCCTGCTGAGAAATGCCCCATGCGCTGAGGT 711
 QY 64 gggaggtgctaccgcccagggcgggaaacaatatgaatcgagttatggcgg 117
 DB 710 GTTAGGTGTTTAGGGCCCAAGGGGAAACCACTTGAGTCTTGTGTGTGG 657

RESULT 9

AAH99496
 ID AAH99496 standard; cDNA; 5654 BP.

XX AAH99496;

DT 16-OCT-2001 (first entry)

DE Human protein encoding cDNA sequence SEQ ID NO:331.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; neurotropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder; ss.

OS Homo sapiens.

XX WO200153455-A2.

PN 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYPSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;
 XX WPI: 2001-457603/49.
 DR P-PSDB; AAM25555.
 XX Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 XX Claim 1; Page 444-446; 1217pp; English.
 XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; neurotropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 XX Sequence 5654 BP; 1078 A; 1631 C; 1682 G; 1263 T; 0 other;

Query Match 8.5%; Score 32.4; DB 22; Length 5654;
 Best Local Similarity 55.3%; Pred. No. 1.1;
 Matches 63; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
 QY 4 cccgcccgcacgcgtatgctttacatcttataatctcctccagggccgagcct 63
 DB 4887 ccagagagcctcccatgctccttgctgtagaattgcccctccatgcgctgaggt 4946
 QY 64 gggaggtgctaccgcccagggcgggaaacaatatgaatcgagttatggcgg 117
 DB 4947 gttaggtggttagggcccaagggaaccacctgagcttctgtgtgtgtgg 5000

RESULT 10

ABA08796

ID ABA08796 standard; cDNA; 5665 BP.

XX ABA08796;

DT 11-JAN-2002 (first entry)

DE Human dJ889M15.3 homologue-encoding cDNA, SEQ ID NO:572.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cytostatic; osteoparic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnerary; antiulcer; ss.

Qy 352 cccgcggctcccccaacctctgctaa 377
||| ||| | |||||
Db 79 TGCCTCCCTTTACACTGCAGTGCTCA 104

RESULT

```

US-09-054-526B-5
; Sequence 5, Application US/09054526B
; Patent No. 6197550
; GENERAL INFORMATION:
; APPLICANT: H TTEN, GERTRUD
; APPLICANT: NEIDHARDT, HELGE
; APPLICANT: BECHTOLD, ROLF
; APPLICANT: POHL, JENS
; TITLE OF INVENTION: DNA SEQUENCES ENCODING NOVEL
; TITLE OF INVENTION: GROWTH/DIFFERENTIATION FACTORS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAIIDO, MARMLSTEIN, MURRAY & ORAM LLP
; STREET: 655 FIFTEENTH STREET, N. W., G STREET LOBBY,
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/054-526B
; FILING DATE: 03-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,222
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 23 190.3
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EPO 9210324.8
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/00350
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KITTS, MONICA CHIN
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P564-8005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from mRNA
; US-09-054-526B-5

```

Query Match 7.9%; Score 30; DB 4; Length 265;
Best Local Similarity 59.3%; Pred. No. 0.14;
Matches 51: Conservative 0; Mismatches 35; Indels 0; Gaps 0;

[illegible]

Db 79 TGCCTCCTTTCACACTGCAGTGCTCA 104

RESULT 6

```

US-09-351-414-3/C
; Sequence 3, Application US/09351414
; Patent No. 6265199
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Baidur, Nand
; APPLICANT: Deisher, Theresa A.
; APPLICANT: Bishop, Paul D.
; FILE OF INVENTION: DISINTEGRIN HOMOLOG
; FILE REFERENCE: 98-29
; CURRENT APPLICATION NUMBER: US/09/351,414
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 2088
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: zintl amino acid deg
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(2088)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2088)
; OTHER INFORMATION: n = A,T,C or G
US-09-351-414-3

```

Query Match 7.9%; Score 30; DB 4; Length 2088;
Best Local Similarity 27.9%; Pred. No. 0.45;
Matches 46: Conservative 36; Mismatches 83; Indels

Qy	165	gcttgcaagtgcagcaaacacttccacgkcccaactcccaagtgagtagctgctggaga	224
		: : : : : : : : : : : : : : : : : : : :	
Db	190	GYTTCNCRTTTCRTARTGDATVTCNACRTARTCSNSWSNARNRNCCTRTTTNARDA	131
Qy	225	agaaagccgcgaagcctctatctaccgtccactgtttttgcgcgagtcggacgcgctagtt	284
Db	130	TNARPTCNARDATRAAYTTNSNNCCRAANGCYTCDATYTGRAANSWNGCYTGNGCNARRT	71
Qy	285	gatctggcgccaaaacactctgtaagtcgccgtgctaagataaa	329
Db	70	GNACNGCYTTRTPTGTGTYTTGTGRTGNCNGCYTTNGTNGTTCNA	26

RESULTS

```

US-08-482-577B-1
; Sequence 1, Application US/08482577B
; Patent No. 5807713
;
; GENERAL INFORMATION:
;
; APPLICANT: HOTTEN, GERTRUD
; APPLICANT: NEIDHARDT, HELGE
; APPLICANT: BECHTOLD, ROLF
; APPLICANT: POHL, JENS
;
; TITLE OF INVENTION: DNA SEQUENCES ENCODING NOVEL
; GROWTH/DIFFERENTIATION FACTORS
;
; NUMBER OF SEQUENCES: 49
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAIIDO, MARMELGSTEIN, MURRAY, AND ORAM
; STREET: 655 FIFTEENTH STREET, N.W., G STREET LOBBY,
; SUITE 330
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:

```



```

RESULT 10
US-09-054-526B-2
: Sequence 2, Application US/09054526B
: Patent No. 6197550
:
: GENERAL INFORMATION:
: APPLICANT: H TTEN, GERTRUD
: APPLICANT: NEIDHARDT, HELGE
: APPLICANT: BECHTOLD, ROLF
: APPLICANT: POHL, JENS
:
: TITLE OF INVENTION: DNA SEQUENCES ENCODING NOVEL
: TITLE OF INVENTION: GROWTH/DIFFERENTIATION FACTORS
: NUMBER OF SEQUENCES: 53
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NIKAIIDO, MARMELSTEIN, MURRAY & ORAM LLP
: STREET: 655 FIFTEENTH STREET, N. W., G STREET LOBBY,
: STREET: SUITE 330
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: USA
: ZIP: 20005-5701
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/054,526B
: FILING DATE: 03-APR-1998

```

Qy 98 atgaatcaggttatggcgggccacaataaaagattccatttggataagaacgaatctg 157
||| ||| ||||||| ||| | ||| ||||| ||| ||| |
Db 731 atggccaagcaatggcgaggctctgatcaaaaatttcaaatggagaaagggaagatta 790

Qy 158 ttaacttgctgagtcgacgcaaccactttccac 192
||| | ||| | ||| | ||| | ||| | ||| |
Db 791 ttaagtcagttgagctggtgacctaattaccac 825

RESULT 12

US-09-387-695-1

; Sequence 1, Application US/09387695

; Patent No. 6280990

; GENERAL INFORMATION:

; APPLICANT: May, Earl

; APPLICANT: Van Horn, Stephanie

; APPLICANT: Warren, Patrick V.

; APPLICANT: Warren, Richard L.

; TITLE OF INVENTION: dnaE

; FILE REFERENCE: GM10237

; CURRENT APPLICATION NUMBER: US/09/387,695

; CURRENT FILING DATE: 1999-08-31

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 3129

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

US-09-387-695-1

Query Match 7.3%; Score 27.8; DB 4; Length 3129;

Best Local Similarity 51.2%; Pred. No. 3.7;

Matches 65; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 99 tgaatcgagttatggcgcccaataaaagattccatttggtataagaacgaactctgt 158
||| | ||| | ||| | ||| | ||| | ||| | ||| |

Db 1893 tggatttagtctgggaagccgatatttgcgtcgagctatgggtataaaggatgcctc 1952

Qy 159 tactttgctgagtcgacgaacacatttccacgtcaccaatccaggtgagtagcttgc 218
||| | ||| | ||| | ||| | ||| | ||| | ||| |

Db 1953 tgcctatgatgataggggctctcttattcaaggctccatagaagcaggccatactgc 2012

Qy 219 tggagaa 225
||| |

Db 2013 ggaaaaa 2019

RESULT 13

US-08-976-259-96

; Sequence 96, Application US/08976259

; Patent No. 6316609

; GENERAL INFORMATION:

; APPLICANT: Dillon, Patrick J.

; APPLICANT: Choi, Gil H.

; APPLICANT: Welch, Rodney A.

; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli

; Patent No. 6316609

; NUMBER OF SEQUENCES: 142

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

; STREET: 1100 New York Ave, N.W., Suite 600

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/976,259

; FILING DATE: Herewith

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953

; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 96:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1128 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-08-976-259-96

Query Match 7.2%; Score 27.6; DB 4; Length 1128;

Best Local Similarity 49.3%; Pred. No. 2.5;

Matches 72; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 72 ctaccggccggaagcgccggaacaatatgaatcgagttatggcgcccaataaaag 131
||| | ||| | ||| | ||| | ||| | ||| | ||| |

Db 495 CTTTCAGCGGTTGGCAGGCCAGGTGTTGATTACACTGATCAGTGGCGGTACATTACCGG 554

Qy 132 attocatttggataagaacgaatctgttactttgttgcagtcgacgcaaccatttcca 191
||| | ||| | ||| | ||| | ||| | ||| | ||| |

Db 555 ACTCGATTCCGTTAAGTTTGACGTACCGATCAGGCTGTATCATCATCTGCCACTCTCGC 614

Qy 192 cgtcaccaatccagtgagtagcttg 217
||| | ||| | ||| | ||| | ||| |

Db 615 CTTCCACCATCAGACCGCGAAGAACATG 640

RESULT 14

US-08-078-090-24

; Sequence 24, Application US/08078090

; Patent No. 5739407

; GENERAL INFORMATION:

; APPLICANT: BERGSTROM, SVEN

; APPLICANT: HERNELL, OLLE

; APPLICANT: LOENNERDAL, BO

; APPLICANT: HJALMARSSON, KARIN

; APPLICANT: HANSSON, LENNART

; APPLICANT: TOERNELL, JAN

; APPLICANT: STROEMQVIST, MATS

; TITLE OF INVENTION: HUMAN BETA-CASEIN PROCESS FOR PRODUCING

; TITLE OF INVENTION: IT AND USE THEREOF

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 SEVENTH STREET, N.W.

; CITY: WASHINGTON

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/078,090

; FILING DATE: 19930618

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/DK92/00236

; FILING DATE: 19-AUG-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/DK91/00233

; FILING DATE: 19-AUG-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: COOPER, IVER P.

; REGISTRATION NUMBER: 28,005

REFERENCE/DOCKET NUMBER: BERGSTROM2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)628-5197
TELEFAX: (202)737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 708 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-078-090-24

Query Match 7.2%; Score 27.4; DB 1; Length 708;
Best Local Similarity 55.9%; Pred. No. 2.3;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 12 cccacgcgcgtatcgtttacatcttatgaatctcgtccacggccgaagcctggagggtg 71
Db 232 CCTATCCCTATGGTTTCTTCCACAAACATTTCTGCTCTGCTCAGCCTGCTGTGGTG 291
QY 72 ctaccggccgaagccggggaacaatatgaatc 104
Db 292 CTGCTGTCTCCTCAGCCTGAAATAATGAAGTC 324

RESULT 15
US-08-249-554-1
Sequence 1, Application US/08249554
Patent No. 5506209

GENERAL INFORMATION:
APPLICANT: Mukerji, P.
APPLICANT: Seo, A.
APPLICANT: Anderson, S.
TITLE OF INVENTION: Product for the Inhibition of Infection of Mammalian Cells by
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lonnie R. Drayer
ADDRESSEE: ROSS Products Division
ADDRESSEE: Abbott Laboratories
STREET: 625 Cleveland Avenue
CITY: Columbus
STATE: Ohio
COUNTRY: United States
ZIP: 43215
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb storage (B)COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh System 7.1(D)SOFTWARE: Clarisworks 1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/249,554
FILING DATE: 26-MAY-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA: NO. 5506209 applicable
TELECOMMUNICATION INFORMATION:
TELEPHONE: (614) 624-3774
TELEFAX: (614) 624-3074
TELEX: No. 5506209e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1065 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Cloned cDNA representing the product of a human
MOLECULE TYPE: genomic DNA segment.
DESCRIPTION: Human milk 'eta-casein
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE: Human
ORGANISM: Homo sapiens

STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: Adult
HAPLOTYPE:
TISSUE TYPE: Mammary Gland
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE: Human Mammary Gland
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD: DNA sequencing and restriction analysis
OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the hum
PUBLICATION INFORMATION:
AUTHORS: B. Lonnerdal et al
TITLE: Cloning and Sequencing of a cDNA encoding human
JOURNAL: Federation European Biochemical Society Letters
VOLUME: 269
ISSUE:
PAGES: 153 - 156
DATE: 1990
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-249-554-1

Query Match 7.2%; Score 27.4; DB 1; Length 1065;
Best Local Similarity 55.9%; Pred. No. 2.9;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 12 cccacgcgcgtatcgtttacatcttatgaatctcgtccacggccgaagcctggagggtg 71
Db 211 CCTATCCCTATGGTTTCTTCCACAAACATTTCTGCTCTGCTCAGCCTGCTGTGGTG 270
QY 72 ctaccggccgaagccggggaacaatatgaatc 104
Db 271 CTGCTGTCTCCTCAGCCTGAAATAATGAAGTC 303

Search completed: June 19, 2002, 16:22:15
Job time: 15793 sec

Thu Jun 20 06:56:42 2002

us-09-462-955b-1_copy_611_991.rni

Page 8

	Query Match	9.2%	Score 35;	DB 12;	Length 524;
	Best Local Similarity	50.0%;	Pred. No. 1.1;	Mismatches 65;	Indels 0; Gaps 0;
	Matches 69;	Conservative	4;		
Qy	67	aggtgctaccggccgaagcgcgggaacaataatgaatcgatttatggcgcgccccacaat	126		
Db	59	AGNTACTTCGGATCCCGTTGCCGACACTATCAGCAACAAGCTAGTGTGGACAA	118		
Qy	127	aaagatctcatcttggaagaacgaatctgttactttgttcagtcgcacgaaccaat	186		
Db	119	ACCANATCCAGTTGGAGGAARAAAATCACAGANNRYTBCTGGAAGTGATAGACAC	178		

Db	179	TTTCMGGAATCCCCCA	196
RESULT	3		
AL667528			
LOCUS	AL667528	898 bp	mRNA
DEFINITION	AL667528 directional larval cDNA library	Ciona intestinalis	cDNA
	clone 020ZG05 5', mRNA sequence.		
ACCESSION	AL667528		
VERSION	AL667528.1	GI:18134435	
KEYWORDS	EST.		
SOURCE	Ciona intestinalis.		
ORGANISM	Ciona intestinalis		
	Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;		
	Phlebobranchia; Cionidae; Ciona.		

TITLE	Ciona intestinalis directional larval cDNA library
JOURNAL	Unpublished (2002)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr IMPORTANT: this sequence may contain errors. The Ciona intestinalis library from which the clone was isolated may be contaminated with cDNAs from bacteria or other Eukarya. Directional larval cDNA library originate from Dr.M.Branno, Stazione A.Dohrn, Naples, Italy, and was prepared in pBluescript2SKt.
FEATURES	Location/Qualifiers
SOURCE	1. .898 /organism="Ciona intestinalis" /db_xref="taxon:7719" /clone="0202G05" /clone_lib="directional larval cDNA library" /note="vector: pBluescript2SKt" 289 a 186 c 157 g 265 t 1 others
BASE COUNT	

	Query Match	9.08;	Score 34.4;	DB 9;	Length 898;
	Best Local Similarity	60.99;	Pred. No. 2.2;	Indels	0;
	Matches 56;	Conservative	0;	Mismatches 36;	Gaps
Qy	169 qcagtgcacgcacacactttccagctcaccacatccagctgagtagcttgcggagaagaa	228			
Db	298 GTATACATGAACAACACATTCCAGCTCACCACCGTGTGAGTATTGCTTGTGCAATG	357			
Qy	229 agccgcagcgcctctatctacogtccattgtt	260			

Accession #	LOCUS	CNS000DBS	1101 bp	DNA	linear	GSS 04-JUN-1999
	DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # RAC27114 of RP11-98 library from Drosophila melanogaster (fruit fly)				

fly), genomic survey sequence.
 AL067144
 AL067144.1 GI:4947840
 GSS.
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org/TheBDGP/Drosophila>
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 Location/Qualifiers
 1. 1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR27J14"
 /note="end : T7"

FEATURES

source

BASE COUNT 247 a 230 c 205 g 246 t 173 others
 ORIGIN

Query Match 8.9%; Score 34; DB 12; Length 1101;
 Best Local Similarity 41.7%; Pred. No. 3.3;
 Matches 50; Conservative 23; Mismatches 47; Indels 0; Gaps 0;
 Qy 126 taaagattccatttgataagaacgaatctgttactttgttcgagtcacgcaaccac 185
 :||| | | : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
 Db 1043 YAAAGSCCCTKWRABATWAAACAAATTTTWTGTTTGTGCAATMACAYC 984
 Qy 186 ttccacgtcaccacatccaggtgagtagctgttgagagaagccgcaagctctatc 245
 :||| | | ||| : | : | : ||| : ||| : ||| : ||| : ||| : ||| :
 Db 983 KSTCCCKAAATCTCCCAVAAWNTYKGGGMAAAWAAWAAAGSCCAHAKCCGWMAC 924

RESULT 5

LOCUS BG006561 403 bp mRNA linear EST 24-JAN-2001
 DEFINITION QV4-GN0143-271100-579-c07 GN0143 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG006561
 VERSION BG006561.1 GI:12449867
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 403)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL
MEDLINE
COMMENT

sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=QV4&t2=QV4-GN0143-271100-579-c07&t3=2000-11-27&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 260.

FEATURES

source

1. 403
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="GN0143"
 /dev_stage="Adult"
 /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
 ; Site_2: SmaI; A mini-library was made by cloning
 products derived from ORESTES PCR (U.S. Letters Patent
 application No. 196,716 - Ludwig Institute for Cancer
 Research) profiles into the puc 18 vector. Reverse
 transcription of tissue mRNA and cDNA amplification were
 performed under low stringency conditions."
 Location/Qualifiers
 1. 403
 74 a 139 c 107 g 83 t

BASE COUNT
ORIGIN

Query Match 8.8%; Score 33.6; DB 10; Length 403;
 Best Local Similarity 47.6%; Pred. No. 2.8;
 Matches 99; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
 Qy 31 atcttgaatatctgtccaggccgaaggccgtggaggtgctaccgccgaagccgg 90
 :||| | | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
 Db 193 ACTCTGTATCCGCTGCTGCTGCCCTCCACAGTCTGAGCTTACCGGCTTTAGTCT 252
 Qy 91 gaacatatgaatcaggttatggcgccgcccacataaaagattccattgataagaac 150
 :||| | | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
 Db 253 GGACACTGAACATTTACTTTGGCTGATCTCATATGAAACCGTCTTTAGCACCAGC 312
 Qy 151 gaatctgttactttgttcgagtcacgcaaccactttccacgtcaccacatccaggtgag 210
 :||| | | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
 Db 313 CAGCTCTTGGCAGTCCAGCAGTGGACACAGCCACGACCCGCGTGTGCGAGC 372
 Qy 211 tagcttgctggagaagaagccgcaagc 238
 :||| | | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
 Db 373 ATGCTGACCGGACCCCTAAAGGCTGAACC 400

RESULT 6

LOCUS AI065300 283 bp mRNA linear EST 21-SEP-2000
 DEFINITION TENU2187 T. cruzi epimastigote normalized cDNA Library Trypanosoma
 cruzi cDNA clone 18e16 5', mRNA sequence.
 ACCESSION AI065300
 VERSION AI065300.1 GI:3340707
 KEYWORDS EST.
 SOURCE Trypanosoma cruzi.
 ORGANISM Trypanosoma cruzi
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma; Schizotrypanum.
 1 (bases 1 to 283)
 Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urmenyi
 T.P., Rondinelli, E., Pettersson, U., Andersson, B. and Aslund, L.
 Gene survey of the pathogenic protozoan Trypanosoma cruzi
 Genome Res. 10 (8), 1103-1107 (2000)
 20414748

COMMENT

Other ESTs: TENU1541
 Contact: Aslund L
 Department of Medical Genetics
 Uppsala University
 Biomedical Center, Box 589, S-751 23 Uppsala, Sweden
 Tel: 46 18 471 45 85
 Fax: 46 18 52 68 49
 Email: lena.aslund@medgen.uu.se
 Seq primer: T7
 High quality sequence stop: 284.

FEATURES

source
 Location/Qualifiers
 1..283
 /organism="Trypanosoma cruzi"
 /strain="Cl-Brenner"
 /db_xref="taxon:5693"
 /clone="18e16"
 /clone_lib="T. cruzi epimastigote normalized cDNA library"
 /cell_type="epimastigote"
 /note="cDNA library constructed with oligo dT primed
 epimastigote mRNA and cloned in pT7t18D phagemid with
 modified polylinker (PHARMACIA)"

BASE COUNT

104 a 71 c 51 g 57 t

FEATURES

source
 Location/Qualifiers
 1..1030
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="2310065E04"
 /sex="male"
 /tissue_type="tongue"
 /dev_stage="adult"
 /lab_host="SOLR"
 /note="Site_1: XhoI; Site_2: SstI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN, Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5',
 GAGAGAGAGATCTCGAGTCTTAATAATTAATCCGCCGCCGCC 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5',
 GAGAGAGAGATCTCGAGTCTTAATAATTAATCCGCCGCCGCC 3']. cDNA
 was cleaved with XhoI and SstI." 1 others

Query Match 8.8%; Score 33.4; DB 9; Length 283;

Best Local Similarity 60.4%; Pred. No. 2.8;
 Matches 55; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 81 cgaaggcgggaacaatgaatgaattatggcgcccaataaagaattccattt 140

Db 52 CGAAAGGCGAGCAACACACGCAATCTTGTGCCGACCAACAGAAATTTTCAG 111

Qy 141 ggataagaacgaatctgttacttgcgtgca 171

Db 112 GAAAGAAATATATGATGATACCTCGCGGAA 142

RESULT

BB610652/c
 LOCUS BB610652 1030 bp mRNA linear EST 31-AUG-2001
 DEFINITION BB610652 RIKEN full-length enriched, adult male tongue Mus musculus
 cDNA clone 2310065E04 5', mRNA sequence.

ACCESSION

BB610652

VERSION

BB610652.1 GI:15391619

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1030)

REFERENCE

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
 M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
 Okazaki,Y., Okido,I., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
 D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 Unpublished (2001)

TITLE

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL

Unpublished (2001)

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,
 URL:http://genome-gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shingawa,A., Aizawa
 K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
 Ishii,Y. and Hayashizaki,Y.

Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
 Func. Genomics 2 pre, L72-L86 (2001)

Please visit our web site (<http://genome-gsc.riken.go.jp>) for
 further details.

e mouse tissues.

FEATURES

source
 Location/Qualifiers
 1..1030
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="2310065E04"
 /sex="male"
 /tissue_type="tongue"
 /dev_stage="adult"
 /lab_host="SOLR"
 /note="Site_1: XhoI; Site_2: SstI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN, Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5',
 GAGAGAGAGATCTCGAGTCTTAATAATTAATCCGCCGCCGCC 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5',
 GAGAGAGAGATCTCGAGTCTTAATAATTAATCCGCCGCCGCC 3']. cDNA
 was cleaved with XhoI and SstI." 1 others

BASE COUNT

285 a 231 c 232 g 281 t

ORIGIN

Query Match 8.7%; Score 33.2; DB 9; Length 1030;
 Best Local Similarity 61.6%; Pred. No. 6;
 Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 286 atctggcgccaaaacctctgttaagtcccgctgaagtataaatagccgcggggtag 345

Db 475 AGCTGAAGCCAGCAATCAACTGAGTCTGGTGAATAGTATACATATTCGCGGACATG 416

Qy 346 tattaccgccgcggctcccaacctc 371

Db 415 TATTCCTCGACTGTTCCCGACCCAC 390

RESULT

8

LOCUS

AZ339575/c

DEFINITION

clone UUGC1M0071B12 F, DNA sequence.

ACCESSION

AZ339575.1

KEYWORDS

GSS.

SOURCE

mouse muscle.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AZ339575 644 bp DNA linear GSS 29-SEP-2000
 1M0071B12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0071B12 F, DNA sequence.

REFERENCE
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 644)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,
and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0071 row: B column: 12
Seq primer: CGTTGTAACAGCGAGCCAGT
Class: plasmid ends
High quality sequence stop: 644.

FEATURES
source

Location/Qualifiers
1..644
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGGCM0071B12"
/clone_lib="Mouse 10kb plasmid UGGCM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gll4732114[gb]/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT
ORIGIN

177 a 186 c 100 g 181 t

Query Match 8.7%; Score 33; DB 12; Length 644;
Best Local Similarity 60.7%; Pred. No. 5.6;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 158 ttactttgctgagtcagcagcaaccactttccacgtccacaatccaggtagtagcttg 217
Db 354 TTTCTGTGCACTGAGTGTAGTTATCTGTCTCTTACAAGCTAAGGCATTTCTGG 295
QY 218 ctggagagaagcgcgcaagcctctatct 246
Db 294 CTTTAAATAAGAGCTGCAAGCCTATATCT 266

RESULT
LOCUS

AV763886 9 762 bp mRNA linear EST 19-OCT-2000
DEFINITION AV763886 MDS Homo sapiens cDNA clone MDSEEE03 5', mRNA sequence.
ACCESSION AV763886
VERSION AV763886.1 GI:10921734

KEYWORDS
SOURCE

human.
Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 762)
Gu, J., Zhao, M., Huang, Q., Xu, X., Li, Y., Peng, Y., Song, H., Xiao, H.,
Li, Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G.,
Yang, Y., Gao, G., Zhang, Q., Chen, S., Han, Z. and Chen, Z.

TITLE

JOURNAL

COMMENT

Homo sapiens cDNA MDS clones
Unpublished (2000)
Contact: Zequan Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source

Location/Qualifiers
1..762
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MDSEEE03"
/clone_lib="MDS"
/tissue_type="Bone marrow"
/cell_type="CD34+ hematopoietic stem/progenitor cell"
/lab_host="BM25.8"
/note="Vector: pTriplex2; Site_1: sfiIA; Site_2: sfiIB"
BASE COUNT 203 a 172 c 177 g 209 t 1 others
ORIGIN

Query Match 8.7%; Score 33; DB 9; Length 762;
Best Local Similarity 48.9%; Pred. No. 6.1;
Matches 87; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 136 catttgataagaacgaactgttacttctgctgagtcagcagcaaccatttccagtc 195
Db 471 CAGCTTGTGCGAAGTACCTGATTTGTTATTAGAACTACCTACCTCTGCTGATTC 530
QY 196 caaatccaggtagtagtctgtctggagaagcgcgcaagcctctctatccagtc 255
Db 531 ACCTTCGCTGGATTCAATGGGTATATGCAAACTGCAGCTTTCATTATGGCATTT 590
QY 256 ttgttttgcgagtcgagcgcgtgagtgatctgctgcccacaaacctctgtaagtc 313
Db 591 GTCTTTGTGAACATCGTGTGCACATTTGTTTAAACAAAAAACTCGCCCAATC 648

RESULT
LOCUS

AQ459250 977 bp DNA linear GSS 23-APR-1999
DEFINITION HS 5078.AL.H03.SP6E.RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-654 Col-5 Row=O, DNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 977)
Mahairas, G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.

REFERENCE

AUTHORS

Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel.: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library PCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 654 row: 0 column: 5
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 977.

	i. .977	
source	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone_ref="plate=654 Col-5 Row-O"	
	/clone_lib="RPC1-11 Human Male BAC Library"	
	/sex="male"	
	/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;	
	Male blood DNA was isolated from one randomly chosen donor	
	and partially digested with a combination of EcoRI and	
	EcoRI Methylase. Size selected DNA was cloned into the	
	pBACE3.6 vector at EcoRI sites"	
BASE COUNT	254 a 223 c 159 q 335 t	6 others

	Query Match	8.7%;	Score 33;	DB 12;	Length 977;
	Best Local Similarity	60.7%;	Pred. NO. 6.8;		
	Matches	54; Conservative	0; Mismatches	35; Indels	0; Gaps
Qy	124	aataaaagattccatttggatagaagcaaatctgttactttgcttgacgcgcgaacc	183		
Db	708	AATAAGAAATATAAATGGGCAATAATCCTTCTTGACCTGTTTTCATTGCACGGACT	767		
Qy	184	actttcacgtcaccatccaggtagta	212		
Db	768	ACGTATTTCCTCTCAAATAGTGGATAATA	796		

RESULT 11
BH437599/c
LOCUS
DEFINITION
 BH437599
 BOHAW63TF BOHA Brassica oleracea genomic clone BOHAW63, DNA
 sequence.
 BH437599
 BH437599.1 GI:17623320
 GSS.
 Brassica oleracea.
ORGANISM
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
 1 (bases 1 to 668)
 Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
OTHER_GSS: BOHAW63TR
COMMENT
 Contact: Chris Town

Email: cdtownetigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES	Location/Qualifiers
• source	1..668

```

Class. sheared ends.
Location/Qualifiers
FEATURES
    . source
    /lab_host="DH10B"
    /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with

```

```

/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHAW63"
/clone_lib="BOHA"
/note="Vector: pHOSt; Site1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOSt using BstXI linkers"
168 a 124 c 184 g 192 t
BASE COUNT
ORIGIN

```

	Query Match	8.6%;	Score 32.8;	DB 12;	Length 668;
	Best Local Similarity	52.1%;	Pred. No. 6.7;		
	Matches 73;	Conservative 0;	Mismatches 67;	Indels 0;	Gaps
Qy	123	caataaagattccatttgataaagaacgaatctgttactttgtctgcagtcgacgcaac	182		
Db	644	CTACAAGAGCTGGCACACAGGTAGGCACAAGTTTGTCTTGTCCACGTGCCCAAGAC	585		
Qy	183	cactttccagctcaccaatccaggtgagtagctgtctggagaagaacgcgaagcctct	242		
Db	584	GACCTTTGTCCACATCACCCCATGATAAGAGCTTCTCTGAGGAGCAGTTACTAGAACTCG	525		
Qy	243	atctaccgtccatttggtttt	262		
Db	524	AGCTTCCAAGCATGACTTCT	505		

RESULT	12
AA221432	
LOCUS	659 bp mRNA linear EST 13-FEB-1997
DEFINITION	my22ell.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA clone IMAGE:696620 5' similar to gb:L23769 Mouse microfilament-associated glycoprotein (MOUSE);, mRNA sequence.
ACCESSION	AA221432
VERSION	AA221432.1 GI:1840686
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 659)

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10

```

source
l. 659
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="IMAGE:696620"
/clone_lib="Barstead mouse pooled organs MPLR54"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT73D-Pac (Pharma

```


VERSION AL084760.1 GI:5285900
 KEYWORDS GSS.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 414)
 AUTHORS Salanoubat,M., Choisine,N., Artiguenave,F., Brottier,P., Wincker,P.,
 Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 414)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 FEATURES
 source
 Location/Qualifiers
 1..414
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone_lib="IGF"
 /clone="F9H2"
 /note="end : T7"
 BASE COUNT 153 a 66 c 58 g 137 t
 ORIGIN

Query Match 8.5%; Score 32.4; DB 12; Length 414;
 Best Local Similarity 60.0%; Pred. No. 7.3;
 Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
 QY 95 aatatgaatcgagtattgaggccacacataaaagattccatttgataagaacgaat 154
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 25 AATATGAATAAAGTTTGTACACTTGAATTAATAATATTCCAAAGCGCTTAACTAAT 84
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 155 ctgttacttgcgtgcagtcgcaacca 184
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 85 CTCTTCTATATAATAACAGTACTAAACCA 114

Search completed: June 19, 2002, 14:04:28
 Job time: 7536 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:48:13 ; Search time 5904.86 Seconds
(without alignments)
995.850 Million cell updates/sec

Title: US-09-462-955B-1_COPY_711_991

Perfect score: 281

Sequence: 1 aatcgagttatggcggccg.....cccccaacctctgtaacccc 281

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htgo.inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	191	68.0	1291	6	ARI172770	ARI172770 Sequence
2	191	68.0	1291	14	CPDCG	M29963 Coconut fol
3	37	13.2	210115	9	AL442127	AL442127 Human DNA
4	36.4	13.0	173848	9	AC090043	AC090043 Homo sapi
5	33	11.7	1438	14	IAU06144	U06144 Influenza A
6	32.8	11.7	29176	8	AP001312	AP001312 Arabidops
7	32.8	11.7	39435	3	CEG49F8	Z70206 Caenorhabdi
8	32.4	11.5	3002	1	AF269928	AF269928 Staphyloc
9	32.4	11.5	3002	6	AX145246	AX145246 Sequence
10	32.4	11.5	3295	1	AF269634	AF269634 Staphyloc
11	32.4	11.5	3295	6	AX144954	AX144954 Sequence
12	32.4	11.5	3518	1	AF270074	AF270074 Staphyloc
13	32.4	11.5	3518	6	AX145392	AX145392 Sequence
14	32.4	11.5	3985	1	AF269816	AF269816 Staphyloc
15	32.4	11.5	3985	6	AX145134	AX145134 Sequence
16	32.4	11.5	105805	2	AP003745	AP003745 Oryza sat
17	32.2	11.5	14907	1	AE007142	AE007142 Mycobacte
18	32.2	11.5	58280	1	MTV014	AL021646 Mycobacte
19	32.2	11.5	168835	9	HS652L8	AL031734 Human DNA
20	32	11.4	245282	2	AC093320	AC093320 Mus muscu
21	31.8	11.3	183539	2	AC091354	AC091354 Rattus no
22	31.6	11.2	118994	9	AC010634	AC010634 Homo sapi
23	31.6	11.2	141456	2	AC108092	AC108092 Homo sapi
24	31.6	11.2	159160	2	AC018624	AC018624 Homo sapi
25	31.6	11.2	172516	2	AC096154	AC096154 Rattus no
26	31.6	11.2	176094	9	AC106799	AC106799 Homo sapi
27	31.4	11.2	98637	2	AL365359	AL365359 Homo sapi
28	31.4	11.2	117919	2	AC053470	AC053470 Mus muscu
29	31.4	11.2	127168	9	HS1163J1	AL031588 Human DNA
30	31.4	11.2	143454	2	AC026820	AC026820 Homo sapi
31	31.4	11.2	316704	9	AF239258S3	AF239391 Homo sapi
32	31.2	11.1	1991	8	AF462831	AF462831 Arabidops
33	31.2	11.1	87246	9	AC002539	AC002539 Homo sapi
34	31.2	11.1	95643	8	F6N15	AF069299 Arabidops
35	31.2	11.1	167234	8	AC004708	AC004708 CIC5B11.1
36	31.2	11.1	175644	2	AC106743	AC106743 Homo sapi
37	31.2	11.1	197119	8	ATCHR1V1	AL161471 Arabidops
38	31	11.0	31	6	ARI172771	ARI172771 Sequence
39	31	11.0	75052	2	AC090558	AC090558 Homo sapi
40	31	11.0	166682	9	AC068194	AC068194 Homo sapi
41	31	11.0	167178	2	AC025672	AC025672 Homo sapi
42	31	11.0	169193	2	AC068523	AC068523 Homo sapi
43	31	11.0	172182	2	AC079816	AC079816 Mus muscu
44	31	11.0	174365	2	AL626785	AL626785 Mus muscu
45	31	11.0	174783	2	AC015565	AC015565 Homo sapi

ALIGNMENTS

RESULT 1	ARI172770	ARI172770	1291 bp	DNA	linear	PAT 17-DEC-2001
LOCUS	Sequence 1 from patent US 6303345.					
DEFINITION	ARI172770					
ACCESSION	ARI172770.1	GI:17912261				
VERSION	Unknown.					
KEYWORDS	Unclassified.					
SOURCE	Unclassified.					
ORGANISM	1 (bases 1 to 1291)					
REFERENCE	Rohde,W., Becker,D., Randles,J.W., Hehn,A. and Salamini,F.					
AUTHORS	Use of a virus DNA as promoter					
TITLE	Patent: US 6303345-A 1 16-OCT-2001;					
JOURNAL	Location/Qualifiers					
FEATURES	source					
	1..1291					
	/organism="unknown"					
BASE COUNT	336 a	323 c	332 g	300 t		
ORIGIN						

```

Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aatcgagttatgagcgccacacataaagaattccatttggataagaacgaatctgtta 60
|||||
Db 1101 AATCGAGTTATGGCGGCCACCAATAAAGATTTCATTGGATAGACGAATCTGTTA 1160

Qy 61 ctttgccttgagtgacgcgaacacatttccacgtcccaatccagtgagtgagcttgc 120
|||||
Db 1161 CTTTGCCTTGAGTGCAGCAACACATTTCCACGTCACCAATCAGGTAGTACGTTGCTG 1220

Qy 121 gaaaaaagccgaagcctctatctacgtccatttgcgagtcgagcgagctg 180
|||||
Db 1221 GAAAGAAACCCGACGCTCTATCTACCGTCCATTGTTTTGCGGATCGACGGCTG 1280

Qy 181 agttgatctgg 191
|||||
Db 1281 AGTTGATCTGG 1291

RESULT 2
CFDGC
LOCUS Coconut foliar decay virus, complete genome.
DEFINITION 1291 bp ss-DNA circular VRL 02-AUG-1993
ACCESSION M29963
VERSION M29963.1 GI:323306
KEYWORDS circular; complete genome.
SOURCE Coconut foliar decay virus
ORGANISM Coconut foliar decay virus; Nanovirus.
REFERENCE
AUTHORS Rohde,W., Randles,J.W., Langridge,P. and Hanold,D.
TITLE Nucleotide sequence of a circular single-stranded DNA associated
JOURNAL with coconut foliar decay virus
MEDLINE Virology 176, 648-651 (1990)
COMMENT Draft entry and printed sequence for [1] kindly submitted by
W.Rhode, 15-MAR-1989, for release after publication.
FEATURES
source
1..1291
Location/Qualifiers
/organism="Coconut foliar decay virus"
/db_xref="taxon:12474"
40..70
/note="stem-loop structure"
103..975
/note="ORF 1"
/codon_start=1
/protein_id="AAA42894.1"
/db_xref="GI:323307"
/translation="MGSSIRRCFTLNYETEEAANVVRRIESLNLVYAIVGDEVAPS
TCGRHLQGFHUKTRRGLKRLVGNDRILHLEPTRGSDQNDRDYCKSERVLLBHVGP
TRPGVKRRLAQRFAEPEDELKLEDPGGYRRCVHVHGSVETWRWAAENPFPPYHNWQ
LEVLSAIGEPADDRITILWICGRDGGDKSVFAKYLGLKPDWFTYTCGTRKDVLYQIE
DKRNLIIDVPRCNLEYLNLYALLLECVKNRAFSSDKYEPLSYLGFHDHVLVFNVLDP
YLIKISDRDKLWNI"
314..775
/note="ORF 2"
/codon_start=1
/protein_id="AAA42895.1"
/db_xref="GI:323308"
/translation="MTGTTWSPVVPVPTNRITETVRRNGCFSTESRLVLESGKHDWPN
DLRLNLSANKTQADTDEALYTELRWNGQGPLKIRSHFHTITGSLKCLLRSSQRT
IAQSSYADETETSGLPNISDSSPTGSHVVEPTCTSTSTRTONEI"
complement(422..568)
/note="ORF 6"
/codon_start=1
/protein_id="AAA42896.1"
/db_xref="GI:323309"
/translation="MEMGTDFQRPILSPPLKRVQIRFGRILGLPGGVHVPQQLVGP
IVAF"
639..797
/note="ORF 3"
/codon_start=1
/protein_id="AAA42897.1"

```

```

/db_xref="GI:323310"
/translation="MRTRRRRREVRCVQISRTQARLVLMHWNQKGRIVPHRGPKTK
FNPRCTQV"
complement(823..987)
/note="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="MTHTLNIPQFYSVPADFQIIRDIGKYEYMHMVEPKITKGFVY
RTECPVLNTP"
1098..1286
/note="ORF 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRVMGGPTIKDSIWIRTNLLCLCQTPQLSTSPIQVSLLEKK
AASLYLPSICFAIGRLS"

BASE COUNT 336 a 323 c 332 g 300 t
ORIGIN

Query Match 68.0%; Score 191; DB 14; Length 1291;
Best Local Similarity 100.0%; Pred. No. 5.5e-51;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aatcgagttatgagcgccacacataaagaattccatttggataagaacgaatctgtta 60
|||||
Db 1101 AATCGAGTTATGGCGGCCACCAATAAAGATTTCATTGGATAGACGAATCTGTTA 1160

Qy 61 ctttgccttgagtgacgcgaacacatttccacgtcccaatccagtgagtgagcttgc 120
|||||
Db 1161 CTTTGCCTTGAGTGCAGCAACACATTTCCACGTCACCAATCAGGTAGTACGTTGCTG 1220

Qy 121 gaaaaaagccgaagcctctatctacgtccatttgcgagtcgagcgagctg 180
|||||
Db 1221 GAAAGAAACCCGACGCTCTATCTACCGTCCATTGTTTTGCGGATCGACGGCTG 1280

Qy 181 agttgatctgg 191
|||||
Db 1281 AGTTGATCTGG 1291

RESULT 3
AL442127 Human DNA sequence from clone RP11-2971f6 on chromosome 13, complete
LOCUS sequence.
DEFINITION AL442127.7 GI:12584450
ACCESSION
VERSION AL442127
KEYWORDS human.
SOURCE HTG.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 210115)
Tromans,A.
Direct Submission
Submitted (26-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humqueres@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jan 28, 2001 this sequence version replaced gi:12578234.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession

```

numbers given in the feature table with their source databases:
 Em: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
 RP11-29716 is from the library RPCI-11.2 constructed by the group
 of Pieter de Jong. For further details see
<http://www.Chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-29716 The true
 left end of clone RP11-207D10 is at 130589 in this sequence. The
 true right end of clone RP11-272L14 is at 88495 in this sequence.

FEATURES

source	Location/Qualifiers
	1..210115
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="13"
	/clone="RP11-29716"
	/clone_lib="RPCI-11.2"
	131..1822
	/note="CpG island"
	/evidence="not_experimental"
	2357..2504
repeat_region	/note="37 copies 4 mer atat 75% conserved"
repeat_region	2363..2502
repeat_region	/note="70 copies 2 mer at 77% conserved"
repeat_region	3453..3755
repeat_region	/note="AluY repeat: matches 1..303 of consensus"
repeat_region	5180..5450
repeat_region	/note="AluSc repeat: matches 1..270 of consensus"
repeat_region	5939..5978
repeat_region	/note="10 copies 4 mer tgga 100% conserved"
repeat_region	9204..9248
repeat_region	/note="15 copies 3 mer gtt 75% conserved"
repeat_region	10198..10397
repeat_region	/note="MER20 repeat: matches 7..218 of consensus"
repeat_region	12023..12404
repeat_region	/note="MSTRD repeat: matches 1..394 of consensus"
repeat_region	13891..14196
repeat_region	/note="AluY repeat: matches 1..302 of consensus"
repeat_region	15743..16019
repeat_region	/note="MLTID repeat: matches 190..505 of consensus"
repeat_region	16034..16726
repeat_region	/note="MLTID repeat: matches 42..880 of consensus"
repeat_region	16727..17162
repeat_region	/note="MLTID repeat: matches 40..482 of consensus"
repeat_region	19077..19259
repeat_region	/note="L2 repeat: matches 2335..2504 of consensus"
repeat_region	19751..20050
repeat_region	/note="AluSg repeat: matches 1..302 of consensus"
repeat_region	20222..20521
repeat_region	/note="AluSg repeat: matches 1..297 of consensus"
repeat_region	25657..25831
repeat_region	/note="Charlie5 repeat: matches 2373..2585 of consensus"
repeat_region	26571..26606
repeat_region	/note="9 copies 4 mer acac 100% conserved"
repeat_region	27856..28166
repeat_region	/note="AluY repeat: matches 1..311 of consensus"
repeat_region	28333..28578
repeat_region	/note="MIR repeat: matches 7..260 of consensus"
repeat_region	28668..28789
repeat_region	/note="AluJb repeat: matches 1..124 of consensus"
repeat_region	29414..29579
repeat_region	/note="MIR repeat: matches 66..259 of consensus"
repeat_region	29712..29855
repeat_region	/note="L2 repeat: matches 1508..1650 of consensus"
repeat_region	30170..30368
repeat_region	/note="L2 repeat: matches 2056..2275 of consensus"
repeat_region	30667..30744
repeat_region	/note="L2 repeat: matches 2672..2749 of consensus"
repeat_region	32012..32087
repeat_region	/note="L1ME2 repeat: matches 6058..6131 of consensus"
misc_feature	32363..33462
misc_feature	/note="CpG island"
repeat_region	/evidence="not_experimental"
repeat_region	32558..32703
repeat_region	/note="73 copies 2 mer cc 63% conserved"
repeat_region	32623..32754
repeat_region	/note="Sequence from overlapping clone BA272L14 (AL138689): Assembly confirmed by restriction digest."
repeat_region	33621..33682
repeat_region	/note="MIR repeat: matches 85..148 of consensus"
repeat_region	33747..33968
repeat_region	/note="MIR repeat: matches 15..257 of consensus"
repeat_region	34151..34515
repeat_region	/note="L1M4 repeat: matches 4316..4702 of consensus"
repeat_region	34520..34746
repeat_region	/note="MLT1A repeat: matches 143..365 of consensus"
repeat_region	34756..34890
repeat_region	/note="MLT1A repeat: matches 3..138 of consensus"
repeat_region	34891..35118
repeat_region	/note="MLT1A repeat: matches 162..365 of consensus"
repeat_region	35316..35546
repeat_region	/note="MLT1A repeat: matches 133..365 of consensus"
repeat_region	35547..35633
repeat_region	/note="MLT1A repeat: matches 63..138 of consensus"
repeat_region	35900..36016
repeat_region	/note="MLT1A repeat: matches 1..117 of consensus"
repeat_region	36404..36604
repeat_region	/note="MLTIC repeat: matches 264..466 of consensus"
repeat_region	36715..37003
repeat_region	/note="AluSg1 repeat: matches 1..289 of consensus"
repeat_region	37056..37085
repeat_region	/note="15 copies 2 mer aa 93% conserved"
repeat_region	37616..37989
repeat_region	/note="THEIC repeat: matches 1..371 of consensus"
repeat_region	39265..39720
repeat_region	/note="MER92B repeat: matches 147..626 of consensus"
repeat_region	39732..40157
repeat_region	/note="THEIC repeat: matches 1..426 of consensus"
repeat_region	40158..40295
repeat_region	/note="MER92B repeat: matches 1..148 of consensus"
repeat_region	40338..42198
repeat_region	/note="L1MCB repeat: matches 115..2032 of consensus"
repeat_region	42660..42698
repeat_region	/note="MERSA repeat: matches 146..187 of consensus"
repeat_region	42996..43434
repeat_region	/note="MLTIC repeat: matches 1..497 of consensus"
repeat_region	44101..44184
repeat_region	/note="MLT1H repeat: matches 448..531 of consensus"
repeat_region	44362..44400
repeat_region	/note="MLT1G repeat: matches 231..269 of consensus"
repeat_region	44516..44571
repeat_region	/note="MLT1H repeat: matches 72..127 of consensus"
repeat_region	45161..45458
repeat_region	/note="AluSg repeat: matches 1..298 of consensus"
repeat_region	45700..45897
repeat_region	/note="MIR repeat: matches 28..236 of consensus"
repeat_region	46028..46270
repeat_region	/note="MER45C repeat: matches 1..228 of consensus"
repeat_region	46466..46695
repeat_region	/note="MER45C repeat: matches 714..952 of consensus"
repeat_region	47322..47737
repeat_region	/note="MSTC repeat: matches 1..405 of consensus"
repeat_region	48351..48562
repeat_region	/note="MLT1J repeat: matches 300..497 of consensus"
repeat_region	48579..49021
repeat_region	/note="L1MC4 repeat: matches 7494..7944 of consensus"
repeat_region	49719..50097
repeat_region	/note="THEIC repeat: matches 3..371 of consensus"
repeat_region	50098..51657
repeat_region	/note="THEIC-internal repeat: matches 1..1580 of consensus"


```

/product="CLC-d chloride channel; anion channel protein"
/protein_id="BAB01934.1"
/db_xref="GI:9294082"
/translation="MVEDNLNIGNSNYNGEGDPESNTLNOLPVKANRTLSSTPLA
LVKAKVSHIESLDEIYENNDLFKIDWRKSKAQVLYVFLKWLACLVGLFTGLIATL
INLAENAVIYKLLAVGHTFQERYVTVGLMVLVGNLGLTLVAVLCVFCFAPTAAGPG
IPEIKAVINGVDYTFMFGATTMIYKIVGSGAVAAAGDGLKEGPLVHGSCIASLLGQ
GTDNHRKRWLRVFNDRDLITCGSAAAGCAAFRSPGVGLFALIEVATWRS
ALLWTFSTVASVLLHREFTEICNSKCGLFGKGLIMDFDVSHVTVYHVTIDIPVM
LIGVIGGLSGLVHLKHLVRLYNLNEKGIHKVLLSLVSLFTSVCLYGLPEFLAK
CKPCDSIDEICPTNGRSGNKQFCHPKGYNDLALLTTNDNDAVRLFSSWTPNEF
GMGSLIFFVYILGFTGATPSGLFPIILMGAAVGRMLGAAGSGSTSIDQGLY
AVLGAALMAGSMRTVSLVIFLETLNLLLLPITMIVLLIARTGDSFNPSYDII
LHLKGLPLEANPEPWRNLTVGELGDAKPPVTLQGVKYSNIVDLVKNTHNAFV
LDEAPVQVGLATGATELHGLLRAHLVKVLLKRWFLTEKRRTEEWVRKGFWDELA
ERENFDVAITSAMEMYVDLHPLTNTPTVTVMENSVAKALVFRQVGLRHLLIIVP
KIQASGMCPVVGILTRDRLAYNIQAPFLLEKSKGKTH"
complement(join(20497..20704,20817..20950,21039..21380,
21519..21632,21759..21842,21918..21997,22092..22146,
22231..22298,22378..22422,22498..22627,22716..22830,
22944..23104))
/note="embI(CAB12631.1
gene_id:MYF5.5
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01935.1"
/db_xref="GI:9294083"
/translation="MNSSLSFSGRFPPIINCRVSAASAPAFVASVPLPSTSDSLA
CSLQPHFQSCGCTQENLHRPAVVDASDFRKYGVDDFTFSCRLWKRCKAKLA
VRGSDNALIGLYQGVTHVVDIPECKNFCDAHPNINAAIELREGIKVDFVFPF
DEDQGTGLRVQMAVTHSTNLRAPERYKNGKQVSLVWNRNERNKADQLST
YLWRKGPNKSFHLIHSVYVNFOTSTNIIFCNWRHLLGERDFWEHVGIDISLPS
SEGOATRAFSLDKLHKYVFGSSVADVAGVIGLSLTSKRCROELRTSVL
NLKRLMYFILMKRLTYFILKSDPSVVKIEVNEARLSFEKTIQRLPNSLCSISWH
HADSVNPLSLIGSDVVDPPRRGLDLSRLQMLESPVSEKRMRSQSSSLNAKE
EKRPWLRAKELSLIQAGNKQTSSENNLTQRLIISCGWESFKEVLENTGKOCCKP
RSF"
complement(24625..24880)
/note="gene_id:MYF5.6
unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
BASE COUNT 9717 a 5015 c 4729 g 9715 t
ORIGIN

```

```

Query Match 11.7%; Score 32.8; DB 8; Length 29176;
Best Local Similarity 52.1%; Pred. No. 17;
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 24 aataaagatttcatttgataaagcaaatctgttactttgcttcagtcagcagcaacc 83
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23766 AATCAACAATTCATTTTGGCAAGCATAGATAGAAATTTTGTTCACACTTTGTATCT 23707
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 84 actttccacgtcccaatccaggtgagtagcttgcgtggagagaagacgcagcctcta 143
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23706 TGTAACTATTATTACCTTCCTCTTTGATTCACAGTTTGAGAGGAACAAGAGAGACTATA 23647
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 144 tctacogtccatttggttttt 163
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 23646 TATGCAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 23627
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
CEC49F8/c CEC49F8 39435 bp DNA linear INV 24-JAN-2002
LOCUS Caenorhabditis elegans cosmid C49F8, complete sequence.
DEFINITION
ACCESSION Z70206
VERSION Z70206.1 GI:1229047
KEYWORDS HFG; monocarboxylate transporter like.
SOURCE Caenorhabditis elegans.

```

```

ORGANISM Caenorhabditis elegans
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
AUTHORS Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
TITLE 1 (sites)
JOURNAL none.
MEDLINE Genome sequence of the nematode C. elegans: a platform for
REMARK Investigating biology. The C. elegans Sequencing Consortium
AUTHORS Science 282 (5396), 2012-2018 (1998)
TITLE 99069613
JOURNAL The C. elegans Sequencing Consortium.
COMMENT 2 (bases 1 to 39435)
DIRECT SUBMISSION Hembray, C.
SUBMITTED (16-MAR-1996) Nematode Sequencing Project, Sanger
INSTITUTE, Hinxton, Cambridge CB10 1SA, England and Department of
GENETICS, Washington University, St. Louis, MO 63110, USA. E-mail:
jesesander.ac.uk or rw@nematode.wustl.edu
CODING SEQUENCES below are predicted from computer analysis, using
PREDICTIONS from Genefinder (P. Green, U. Washington), and other
AVAILABLE INFORMATION.
CURRENT SEQUENCE finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
EXCEPTIONS are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone C49F8.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone C49F8 is at 1 in this sequence. The true
left end of clone F02C12 is at 39332 in this sequence. The start
right end of clone F46G10 is at 1490 in this sequence. The start of
this sequence (1..1490) overlaps with the end of sequence Z50177.
The end of this sequence (39332..39435) overlaps with the start of
sequence Z54269.
For a graphical representation of this sequence and its analysis
see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
name=C49F8
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
FEATURES
Location/Qualifiers
source 1..39435
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="C49F8"
complement(join(1006..1201,1257..1401,1454..1727,
1779..1871))
/gene="F46G10.6"
complement(join(1006..1201,1257..1401,1454..1727,
1779..1871))
/gene="F46G10.6"
/translation="MSAIVEDMYLSSSMKWEKQFRKRHHSDSDSSSPKSPKSPSPM
DDRRHNELEERRRRHHIKDHTILKDAIPLLDGKSSPALILKRAVEFIHYMOTKL
SSQKATIEDLTRKNLEERLEERSSGSSSSRLPALAVSSQMQLTPIIPIQMONI
AQLSYQQQANIQAQSTPAQLDGLIALNDAILLGSFQSPSLSDSAPGTFFSG
FYPCAFSPVDQQAQVKT"
complement(join(12393..12614,12767..12955,13353..13460,
13504..13556,14826..14866,15183..15266,15849..16375))
/gene="C49F8.1"
complement(join(12393..12614,12767..12955,13353..13460,
13504..13556,14826..14866,15183..15266,15849..16375))
/gene="C49F8.1"
/note="predicted using Genefinder"

```


[illegible]


```

Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE
AUTHORS
  Kimmerly, W.J., Taylor, J. David., Nelsen, A.J., Godlevski, M.M.,
  Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
  Listenebee, S., Ashanti, C., Altshuller, G., Mamo, L., Shepherd, N.S.,
  Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
  Furdon, P.J.
TITLE
  Transposon-mediated sequencing of the Staphylococcus epidermidis
  genome
JOURNAL
  Unpublished
REFERENCE
AUTHORS
  Taylor, J. David., Kimmerly, W.J., Nelsen, A.J., Godlevski, M.M.,
  Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
  Listenebee, S., Ashanti, C., Altshuller, G., Mamo, L., Shepherd, N.S.,
  Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
  Furdon, P.J.
TITLE
  Direct Submission
JOURNAL
  Submitted (22-MAY-2000) Departments of Genomic Sciences and
  Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
  Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES
  source
    1. .3518
      /organism="Staphylococcus epidermidis"
      /strain="SRI"
      /db_xref="taxon:1282"
      /clone="step.1047b09"
BASE COUNT
  1091 a 677 c 495 g 1255 t
ORIGIN
  1091 atccaggtgagtagctgtggaagaagccgcgaagcctctctatccgtccattgt 159
  2590 atttaagtgtattcttagtatagaaatgacctgagacatgattgcccaggcatttta 2649
  160 ttgtgcgcatcgacgcgtgagttgattctgtggcccaaaacctctctgaagtcctg 219
  2650 tttattggcaattgcacacccagtgtagtgaactgggctaattctcaattttcaactctagtc 2709
  220 taagtataaa 229
  2710 TCATTACCAA 2719

Query Match 11.5%; Score 32.4; DB 1; Length 3518;
Best Local Similarity 53.1%; Pred. No. 20;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 100 atccaggtgagtagctgtggaagaagccgcgaagcctctctatccgtccattgt 159
Db 2590 atttaagtgtattcttagtatagaaatgacctgagacatgattgcccaggcatttta 2649
Qy 160 ttgtgcgcatcgacgcgtgagttgattctgtggcccaaaacctctctgaagtcctg 219
Db 2650 tttattggcaattgcacacccagtgtagtgaactgggctaattctcaattttcaactctagtc 2709
Qy 220 taagtataaa 229
Db 2710 TCATTACCAA 2719

RESULT 13
AX145392
LOCUS
  AX145392 Sequence 4114 from Patent WO0134809. linear PAT 31-MAY-2001
DEFINITION
  AX145392
ACCESSION
  AX145392.1 GI:14283957
VERSION
  AX145392.1
KEYWORDS
  synthetic construct.
  synthetic construct
  artificial sequence.
  1 (bases 1 to 3518)
  Kimmerly, W.J.
ORGANISM
  Staphylococcus epidermidis nucleic acids and proteins
REFERENCE
  Patent: WO 0134809-A 4114 17-MAY-2001;
  GLAXO GROUP LIMITED (GB)
  Location/Qualifiers
    1. .3518
      /organism="synthetic construct"
      /db_xref="taxon:32630"
      /note="synthetic nucleic acid sequence"
BASE COUNT
  1091 a 677 c 495 g 1255 t
ORIGIN
  1091 atccaggtgagtagctgtggaagaagccgcgaagcctctctatccgtccattgt 159
  2590 atttaagtgtattcttagtatagaaatgacctgagacatgattgcccaggcatttta 2649
  160 ttgtgcgcatcgacgcgtgagttgattctgtggcccaaaacctctctgaagtcctg 219
  2650 tttattggcaattgcacacccagtgtagtgaactgggctaattctcaattttcaactctagtc 2709
  220 taagtataaa 229
  2710 TCATTACCAA 2719

Query Match 11.5%; Score 32.4; DB 6; Length 3518;

```

```

Best Local Similarity 53.1%; Pred. No. 20;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 100 atccaggtgagtagctgtggaagaagccgcgaagcctctctatccgtccattgt 159
Db 2590 atttaagtgtattcttagtatagaaatgacctgagacatgattgcccaggcatttta 2649
Qy 160 ttgtgcgcatcgacgcgtgagttgattctgtggcccaaaacctctctgaagtcctg 219
Db 2650 tttattggcaattgcacacccagtgtagtgaactgggctaattctcaattttcaactctagtc 2709
Qy 220 taagtataaa 229
Db 2710 TCATTACCAA 2719

RESULT 14
AF269816
LOCUS
  AF269816 Staphylococcus epidermidis strain SRI clone step.1024b10 genomic
  sequence. linear BCT 01-AUG-2000
DEFINITION
  AF269816
ACCESSION
  AF269816.1 GI:9623714
VERSION
  AF269816.1
KEYWORDS
  Staphylococcus epidermidis.
  Staphylococcus epidermidis
  Bacteria; Firmicutes; Bacillus/Clostridium group;
  Bacillus/Staphylococcus group; Staphylococcus.
  1 (bases 1 to 3985)
  Kimmerly, W.J., Taylor, J. David., Nelsen, A.J., Godlevski, M.M.,
  Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
  Listenebee, S., Ashanti, C., Altshuller, G., Mamo, L., Shepherd, N.S.,
  Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
  Furdon, P.J.
TITLE
  Transposon-mediated sequencing of the Staphylococcus epidermidis
  genome
JOURNAL
  Unpublished
REFERENCE
  2 (bases 1 to 3985)
  Taylor, J. David., Kimmerly, W.J., Nelsen, A.J., Godlevski, M.M.,
  Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I.,
  Listenebee, S., Ashanti, C., Altshuller, G., Mamo, L., Shepherd, N.S.,
  Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and
  Furdon, P.J.
TITLE
  Direct Submission
JOURNAL
  Submitted (22-MAY-2000) Departments of Genomic Sciences and
  Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
  Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES
  source
    1. .3985
      /organism="Staphylococcus epidermidis"
      /strain="SRI"
      /db_xref="taxon:1282"
      /clone="step.1024b10"
BASE COUNT
  1175 a 779 c 583 g 1448 t
ORIGIN
  1175 atccaggtgagtagctgtggaagaagccgcgaagcctctctatccgtccattgt 159
  381 atttaagtgtattcttagtatagaaatgacctgagacatgattgcccaggcatttta 440
  160 ttgtgcgcatcgacgcgtgagttgattctgtggcccaaaacctctctgaagtcctg 219
  441 tttattggcaattgcacacccagtgtagtgaactgggctaattctcaattttcaactctagtc 500
  220 taagtataaa 229
  501 TCATTACCAA 510

Query Match 11.5%; Score 32.4; DB 1; Length 3985;
Best Local Similarity 53.1%; Pred. No. 20;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 100 atccaggtgagtagctgtggaagaagccgcgaagcctctctatccgtccattgt 159
Db 381 atttaagtgtattcttagtatagaaatgacctgagacatgattgcccaggcatttta 440
Qy 160 ttgtgcgcatcgacgcgtgagttgattctgtggcccaaaacctctctgaagtcctg 219
Db 441 tttattggcaattgcacacccagtgtagtgaactgggctaattctcaattttcaactctagtc 500
Qy 220 taagtataaa 229
Db 501 TCATTACCAA 510

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 17:05:17 ; Search time 3798.65 Seconds
(without alignments)
127.006 Million cell updates/sec

Title: US-09-462-955B-1_COPY_711_991
Perfect score: 281
Sequence: 1 aatgagtgatggcgccgccc.....cccccaactctgttaacccc 281

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /net/abs06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	32.4	11.5	3002	AAH54604	S. epidermidis gen
2	32.4	11.5	3295	AAH54312	S. epidermidis gen
3	32.4	11.5	3518	AAH54750	S. epidermidis gen
4	32.4	11.5	3985	AAH54492	S. epidermidis gen
5	32.2	11.5	4411529	AA199682	Mycobacterium tube
6	31.4	11.2	460	AA196825	Human polynucleoti
7	31	11.0	31	AAQ56057	Coconut Foliar Dec
8	31	11.0	31	AAQ02688	Gemini virus DNA f
9	31	11.0	31	AAQ02686	Gemini virus DNA f

c	10	30.4	10.8	974	21	AA666416	Human secreted pro
c	11	30	10.7	265	14	AAQ47710	TGF-beta-like clon
c	12	30	10.7	2088	21	AAZ45827	Degenerate DNA enc
c	13	30	10.7	2088	22	AA511992	Human degenerate D
c	14	30	10.7	2272	17	AA111104	Transferring gp1
c	15	30	10.7	2272	22	AA744421	Human TGF-beta MPT
c	16	29.8	10.6	1636	22	AAH77087	Human hepatic nucl
c	17	29.6	10.5	30600	20	AAZ32023	Human METH1 relate
c	18	29.6	10.5	30601	22	AA60080	259361 cDNA clone.
c	19	29.4	10.5	400	18	AAV78089	Staphylococcus aur
c	20	29.4	10.5	1194	23	AA54726	Staphylococcus aur
c	21	29.4	10.5	14078	18	AAV74502	Staphylococcus aur
c	22	29.4	10.5	48452	23	ABL07108	Drosophila melanog
c	23	29.2	10.4	8877	23	AA57138	DNA encoding Droso
c	24	29.2	10.4	8877	23	ABL13920	Drosophila melanog
c	25	28.6	10.2	315	22	ABA18519	Human nervous syst
c	26	28.6	10.2	591	23	AA580984	DNA encoding novel
c	27	28.6	10.2	3031	22	ABA18518	Human nervous syst
c	28	28.6	10.2	5531	22	AAK78665	Human immune/haema
c	29	28.2	10.0	858	22	AA123310	Human breast cance
c	30	28.2	10.0	1414	19	AAV26248	Genomic DNA SEQ ID
c	31	28.2	10.0	1497	21	AAZ50194	Male fusion plasm
c	32	28	10.0	978	23	AA582110	DNA encoding novel
c	33	28	10.0	978	23	AA586806	DNA encoding novel
c	34	28	10.0	1509	23	AA52691	E. coli DNA for ce
c	35	28	10.0	2493	23	ABL26334	Drosophila melanog
c	36	28	10.0	3737	21	AA63739	Maize RNA-directed
c	37	28	10.0	5164	23	ABL01909	Drosophila melanog
c	38	28	10.0	10953	22	AA53689	Genomic sequence #
c	39	28	10.0	10953	22	AAK90032	Human digestive sy
c	40	28	10.0	30598	23	ABL01908	Drosophila melanog
c	41	28	10.0	80450	21	AA62295	BAC containing rep
c	42	27.8	9.9	522	18	AAV75313	Staphylococcus aur
c	43	27.6	9.8	1317	23	ABL13049	Drosophila melanog
c	44	27.6	9.8	1378	14	AAQ42310	RGT precursor. Or
c	45	27.6	9.8	3317	23	ABL13048	Drosophila melanog

ALIGNMENTS

RESULT 1
AAH54604
ID AAH54604 standard; DNA; 3002 BP.
AC AAH54604;
XX
XX
DT 03-SEP-2001 (first entry)
XX
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3968.
XX
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
XX vaccination; endocarditis; ds.
XX
XX
OS Staphylococcus epidermidis.
XX
XX
PN WO200134809-A2.
XX
XX
PD 17-MAY-2001.
XX
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
XX
PR 09-NOV-1999; 99US-0164258.
XX
XX
PA (GLAX) GLAXO GROUP LTD.
XX
XX
PI Kimmerly WJ;
XX
XX
DR WPI; 2001-316495/33.
XX
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -

PS Claim 8; Page 1634-1635; 2180pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides

CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.

CC (I) and (II) can have antibacterial activity and therefore can be used

CC in vaccination. The nucleic acids (I) may be used to produce the

CC S. epidermidis polypeptides (II) via the production of vectors

CC containing them which are used to produce host cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their

CC activity and therefore identify compounds that may be used for the

CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to

CC AAH55090 represent specifically claimed S. epidermidis genomic DNA

CC polynucleotide sequences from the present invention. AAH55091 to

CC AAH55098 represent oligonucleotide sequences and primers which are used

CC in the exemplification of the present invention.

CC N.B. The present invention specifically claims all the polynucleotide

CC sequences given in the sequence listing of the present specification,

CC however the sequence listing only goes up to SEQ ID NO:4454 so even

CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,

CC no sequences are present for SEQ ID NO:4455 to 4464.

XX SQ Sequence 3002 BP; 899 A; 595 C; 418 G; 1090 T; 0 other;

Query Match 11.5%; Score 32.4; DB 22; Length 3002;

Best Local Similarity 53.1%; Pred. No. 0.54;

Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 100 attcaggtagtgagctgctggagaagaagccgcgaagcctctatcaccgtccattgt 159

DB 627 atttaagtattctcttagtatagaaaatgcctgagacaatgattgccaggcatttta 686

QY 160 ttgtgcgcagtcgagcggctgagtgatctgctgcgcgcacaaacctctgtaagtcctgtgc 219

DB 687 ttattggcaattgacaacgcagtggaacttggtggaactctcaatttcaactctagtcc 746

QY 220 taagtataaa 229

DB 747 tcattaccaa 756

RESULT 2

AAH54312

ID AAH54312 standard; DNA; 3295 BP.

XX AAH54312;

AC

DT 03-SEP-2001 (first entry)

XX

DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3676.

XX

KW *Staphylococcus epidermidis* SRI strain; infection; diagnosis;

KW vaccination; endocarditis; ds.

XX

OS *Staphylococcus epidermidis*.

XX

PN WO200134809-A2.

XX

PD 17-MAY-2001.

XX

PF 09-NOV-2000; 2000WO-US30782.

XX

PR 09-NOV-1999; 99US-0164258.

XX

PA (GLAX) GLAXO GROUP LTD.

XX

PI Kimmerly WJ;

XX

PI Kimmerly WJ;

XX

DR WPI; 2001-316495/33.

XX

PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,

useful for vaccinating against infections, e.g. endocarditis -

PT Claim 8; Page 1285-1287; 2180pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides

CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.

CC (I) and (II) can have antibacterial activity and therefore can be used

CC in vaccination. The nucleic acids (I) may be used to produce the

CC S. epidermidis polypeptides (II) via the production of vectors

CC containing them which are used to produce host cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their

CC activity and therefore identify compounds that may be used for the

CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to

CC AAH55090 represent specifically claimed S. epidermidis genomic DNA

CC polynucleotide sequences from the present invention. AAH55091 to

CC AAH55098 represent oligonucleotide sequences and primers which are used

CC in the exemplification of the present invention.

CC N.B. The present invention specifically claims all the polynucleotide

CC sequences given in the sequence listing of the present specification,

CC however the sequence listing only goes up to SEQ ID NO:4454 so even

CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,

CC no sequences are present for SEQ ID NO:4455 to 4464.

XX SQ Sequence 3295 BP; 971 A; 667 C; 456 G; 1201 T; 0 other;

Query Match 11.5%; Score 32.4; DB 22; Length 3295;

Best Local Similarity 53.1%; Pred. No. 0.57;

Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 100 attcaggtagtgagctgctggagaagaagccgcgaagcctctatcaccgtccattgt 159

DB 776 atttaagtattctcttagtatagaaaatgcctgagacaatgattgccaggcatttta 835

QY 160 ttgtgcgcagtcgagcggctgagtgatctgctgcgcgcacaaacctctgtaagtcctgtgc 219

DB 836 ttattggcaattgacaacgcagtggaacttggtggaactctcaatttcaactctagtcc 895

QY 220 taagtataaa 229

DB 896 tcattaccaa 905

RESULT 3

AAH54750

ID AAH54750 standard; DNA; 3518 BP.

XX AAH54750;

AC

DT 03-SEP-2001 (first entry)

XX

DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:4114.

XX

KW *Staphylococcus epidermidis* SRI strain; infection; diagnosis;

KW vaccination; endocarditis; ds.

XX

OS *Staphylococcus epidermidis*.

XX

PN WO200134809-A2.

XX

PD 17-MAY-2001.

XX

PF 09-NOV-2000; 2000WO-US30782.

XX

PR 09-NOV-1999; 99US-0164258.

XX

PA (GLAX) GLAXO GROUP LTD.

XX

PI Kimmerly WJ;

XX

PI Kimmerly WJ;

XX

DR WPI; 2001-316495/33.

XX Fleischmann RD, White OR, Fraser CM, Venter JC;
 XX WPI; 2001-647261/74.
 XX
 XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
 XX determining the nucleotide sequence of the strain at positions in the
 XX genome corresponding to positions where M. tuberculosis strains CDC
 XX 1551 and H37Rv differ
 XX
 XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
 XX
 XX The invention relates to evaluating strain variation within and between
 XX different populations of the tuberculosis bacterial pathogen,
 XX Mycobacterium tuberculosis or related Mycobacterium by determining the
 XX nucleotide sequence of the first strain at positions in the complete
 XX sequence of the genome that correspond to positions that differ in the
 XX nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
 XX H37Rv (AAI99682). The method is useful for evaluating strain variation of
 XX M. tuberculosis and has valuable application in the fields of
 XX tuberculosis genetics, epidemiology, patient treatment and epidemic
 XX monitoring.
 XX Note: The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from USPTO
 XX at seqdata.uspto.gov/sequence.html?docID=6294328B1.
 XX
 XX Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
 XX
 XX
 XX Query Match 11.5%; Score 32.2; DB 22; Length 4411529;
 XX Best Local Similarity 57.4%; Pred. No. 15;
 XX Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 XX
 XX QY 108 gagtagctgtggaagaagcgcgaagcctctatctatcctccattgtttgctgc 167
 XX Db 3577728 GAGCGGCACGGCGGCAACAAACACCGCGAGCTGCCGTGCGGATCGGTGAC 3577669
 XX
 XX QY 168 gatcgagcgcgtgagttgatctggcgcaaaaacctctgt 208
 XX Db 3577668 GCACTGTTGCCGAGCGGCGAGGAGGATCCAACTGCT 3577628
 XX
 XX RESULT 6
 XX AAI86825/c
 XX ID AAI86825 standard; cDNA: 460 BP.
 XX AC AAI86825;
 XX
 XX 06-NOV-2001 (first entry)
 XX
 XX Human polynucleotide SEQ ID NO 6885.
 XX
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 XX tissue growth factor; immunomodulatory; cancer; leukaemia;
 XX nervous system disorders; arthritis; inflammation; ss.
 XX
 XX Homo sapiens.
 XX
 XX WO200164835-A2.
 XX
 XX 07-SEP-2001.
 XX
 XX 26-FEB-2001; 2001WO-US04927.
 XX
 XX 28-FEB-2000; 2000US-0515126.
 XX
 XX 18-MAY-2000; 2000US-0577409.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT;
 XX
 XX WPI; 2001-514838/56.
 XX
 XX

DR P-PSDB; AAO06894.
 XX
 XX Isolated nucleic acids and polypeptides, useful for preventing
 XX diagnosing and treating e.g. leukaemia, inflammation and immune
 XX disorders -
 XX
 XX Claim 1; SEQ ID NO 6885; 1399pp + Sequence Listing; English.
 XX
 XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
 XX the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 XX cytokine, cell proliferation or cell differentiation or which may induce
 XX production of other cytokines in other cell populations. The
 XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
 XX peptide therapy. The polypeptides have various cytokine-like activities,
 XX e.g. stem cell growth factor activity, haematopoiesis regulating
 XX activity, tissue growth factor activity, immunomodulatory activity and
 XX activin/inhibin activity and may be useful in the diagnosis and/or
 XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
 XX inflammation.
 XX Note: The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 460 BP; 114 A; 115 C; 141 G; 88 T; 2 other;
 XX

Query Match 11.2%; Score 31.4; DB 22; Length 460;
 Best Local Similarity 48.1%; Pred. No. 0.55;
 Matches 89; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
 QY 85 ctctccacgtccaccaatccagtgtagctgtctggaagaagcgcgaagcctctat 144
 Db 377 CTGTATCATCTGCCAAGCTGCTCTCAAGCTTTGGCATCAAGCAATCTCGCCCTCAGC 318
 QY 145 ctacgtccattgttttgcgcgacgcgcgtgagttgatctgagcgcacaaacctc 204
 Db 317 CTCCCATCATCTGGGATTACAGGTGCACATCAGCCTATAGGGGTTTTCCTTCAGTTTC 258
 QY 205 tgctaaagtcgcgtctaagataataagcgcgcgggggtagtagtattacccccgcgcctccc 264
 Db 257 TGACTTTGAGGAGGTCATTGGAAACAGACCCCTGGGCGCTGTCTCCCGCTGAGCCCCAC 198
 QY 265 caacc 269
 Db 197 TGCCC 193

RESULT 7
 AAO56057/c
 ID AAO56057 standard; DNA; 31 BP.
 XX
 XX AAO56057;
 XX AC
 XX 12-AUG-1994 (first entry)
 XX DT
 XX Coconut Foliar Decay Virus promoter stem-loop.
 XX DE
 XX Coconut Foliar Decay Virus; CFDV; strong promoter; tissue-specific;
 XX phloem-specific; stem-loop structure; transgenic plant; ds.
 XX KW
 XX Coconut Foliar Decay Virus.
 XX OS
 XX
 XX Key Location/Qualifiers
 XX stem_loop 1..31
 XX /*tag= a
 XX /function= promoter
 XX /note= "loop has homology to geminivirus sequence"
 XX
 XX DE4306832-C.
 XX
 XX 24-FEB-1994.
 XX
 XX 04-MAR-1993; 93DE-4306832.
 XX PF

```

XX PR 04-MAR-1993; 93DE-4306832.
XX PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Becker D, Randles JW, Rohde W, Salamini F;
XX PT WPI; 1994-058406/08.
XX DR Use of coconut foliar decay virus DNA as promoter - for
XX PT tissue-specific gene expression in transgenic plants
XX PS Claim 1; Fig 2; 8pp; German.
XX
XX A DNA fragment from the CFDV genome can be used as a phloem-specific
XX CC promoter in the construction of transgenic plants. The promoter is
XX CC strong; it has 30-50% of the activity of the CamV 35S promoter in
XX CC tobacco plants.
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 11.0%; Score 31; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 agccgcggggctagtagtattaccccgcggt 261
DB 31 AGCCGCGGGGCTAGTATTACCCCGCGGCT 1

RESULT 8
AAAX02688/c
ID AAX02688 standard; DNA; 31 BP.
XX AC AAX02688;
XX DT 10-MAY-1999 (first entry)
XX DE Gemini virus DNA fragment stem loop.
XX KW Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.
XX OS Gemini virus.
XX FH Key Location/Qualifiers
FT stem_loop 1..31 /*tag= a
FT misc_binding 1..11 /*tag= b
FT /*note= "Region binds to nucleotides 21 to 31"
FT misc_binding 21..31 /*tag= c
FT /*note= "Region binds to nucleotides 1 to 11"
XX DE19730502-A1.
XX PN 21-JAN-1999.
XX PD 16-JUL-1997; 97DE-1030502.
XX PF 16-JUL-1997; 97DE-1030502.
XX PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PA Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX PI WPI; 1999-096863/09.
XX DR Coconut foliar decay virus promoters - for gene expression in
XX PT bacteria and yeasts
XX PS Disclosure; Fig 2; 12pp; German.
XX
XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX CC that includes the stem-loop structure of CFDV DNA but lacks the
XX CC translation start codons of open reading frames ORF1 and/or ORF2. The
XX CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX CC promoter for tissue-specific (especially phloem-specific) gene expression
XX CC in plants and for production of chimeric constructs for transient or
XX CC stable expression. Certain fragments of CFDV DNA have stronger promoter
XX CC activity in E. coli than the CamV 35S promoter.
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 11.0%; Score 31; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 agccgcggggctagtagtattaccccgcggt 261
DB 31 AGCCGCGGGGCTAGTATTACCCCGCGGCT 1

RESULT 9
AAAX02686/c
ID AAX02686 standard; DNA; 31 BP.
XX AC AAX02686;
XX DT 10-MAY-1999 (first entry)
XX DE Gemini virus DNA fragment stem loop.
XX KW Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.
XX OS Gemini virus.
XX FH Key Location/Qualifiers
FT stem_loop 1..31 /*tag= a
FT misc_binding 1..11 /*tag= b
FT /*note= "Region binds to nucleotides 21 to 31"
FT misc_binding 21..31 /*tag= c
FT /*note= "Region binds to nucleotides 1 to 11"
XX DE19730535-A1.
XX PN 21-JAN-1999.
XX PD 16-JUL-1997; 97DE-1030535.
XX PF 16-JUL-1997; 97DE-1030535.
XX PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PA Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX PI WPI; 1999-096867/09.
XX DR Coconut foliar decay virus promoters - for gene expression in
XX PT plants, bacteria and yeasts
XX PS Disclosure; Fig 2; 12pp; German.
XX
XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX CC that includes the stem-loop structure of CFDV DNA but lacks the
XX CC translation start codons of open reading frames ORF1 and/or ORF2. The
XX CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX CC promoter for tissue-specific (especially phloem-specific) gene expression
XX CC in plants and for production of chimeric constructs for transient or

```

CC stable expression. Certain fragments of CFV DNA have stronger promoter
 XX activity in E. coli than the CMV 35S promoter.
 SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 11.0%; Score 31; DB 20; Length 31;
 Best Local Similarity 100.0%; Pred. No. 0.23;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 agccgcgggctagattaccccgccgct 261
 |||||
 DB 31 AGCCGCGGGCTAGTATTACCCCGCGCT 1

RESULT 10
 AAC66416/c
 ID AAC66416 standard; DNA; 974 BP.

XX AC AAC66416;

XX DT 14-FEB-2001 (first entry)

XX DE Human secreted protein coding sequence SEQ ID NO: 17.

XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 XX KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 XX KW neurological disease; infection; human; secreted protein; ss.

XX OS Homo sapiens.

XX PN WO200058350-A1.

XX PD 05-OCT-2000.

XX PF 22-MAR-2000; 2000WO-US07483.

XX PR 26-MAR-1999; 99US-0126596.

XX PR 22-DEC-1999; 99US-0171552.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM, Komatsoulis G;

XX DR WPI: 2000-602357/57.

XX DR P-PSDB; AAB32008.

XX PT Nucleic acid molecules encoding human secreted proteins, used in
 XX PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 XX PT Parkinson's diseases and cancers -

XX PS Claim 1; Page 337; 423pp; English.

XX CC The invention relates to the isolation of genes AAC66410-C66458 encoding
 CC 49 human secreted proteins AAB32002-B32050. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (SEQID) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

CC and parasitic infections.

XX

SQ Sequence 974 BP; 229 A; 252 C; 263 G; 230 T; 0 other;

Query Match 10.8%; Score 30.4; DB 21; Length 974;
 Best Local Similarity 54.5%; Pred. No. 1.7;
 Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 138 cctctatctacgcgtccattgttttgcgcgatcgacggctgagttgatctggcgccaa 197
 |||||
 DB 441 CCTCTAAGATCCCTTCAACTCCGAATCTCTGATTTGACTGCGAGGAAACAAAGCCCTAG 382

QY 198 aaacctctgctaaagtcccgctgctaaagtataaaatagccgcgggctagtatt 249
 |||||
 DB 381 AAACATGGGTTACAGCCTGGGCTAGCTTGCTTAGCGGCCAGGCGAGGTTTT 330

RESULT 11

AAQ47710

ID AAQ47710 standard; cDNA to mRNA; 265 BP.

XX AC AAQ47710;

XX DT 11-FEB-1994 (first entry)

XX DE TGF-beta-like clone MP-121.

XX KW Human; transforming growth factor; beta; TGF-beta; pharmaceutical;
 XX KW bone; cartilage; tooth; wound repair; immunosuppressor;
 XX KW organ transplant; cosmetic surgery; antibody; diagnosis; ss.

XX OS Homo sapiens.

XX FH Key

XX FT Location/Qualifiers

XX FT CDS

XX FT 2..265

XX FT /*tag= a

XX PN WO9316099-A.

XX PD 19-AUG-1993.

XX PF 12-FEB-1993; 93WO-EP00350.

XX PR 12-FEB-1992; 92EP-0102324.

XX PA (BIOP-) BIOPHARM GES BIOTECHNOLOGISCHEN ENTWICKL.

XX PI Hoetten G, Neidhardt H;

XX DR WPI: 1993-272824/34.

XX DR P-PSDB; AAR45447.

XX PT New transforming growth factor-beta family proteins and DNA -
 XX PT used in tissue and wound repair, in treatment of bone, cartilage
 XX PT and tooth defects, and antibodies for diagnosis

XX PS Claim 4; Page 18; 29pp; English.

XX CC The sequences given in AAQ47709-10 represent embryo and liver derived
 CC human transforming growth factor-beta (TGF-beta) genes respectively.
 CC The proteins encoded by these sequences may be used in a pharmaceutical
 CC composition for the treatment of various bone, cartilage or tooth
 CC defects and in tissue and wound repair processes. These proteins
 CC may also be used as immunosuppressors in organ transplants and in
 CC cosmetic surgery. Antibodies raised against these proteins may be
 CC used for diagnostic purposes.

SQ Sequence 265 BP; 58 A; 80 C; 69 G; 58 T; 0 other;

Query Match

Best Local Similarity 10.7%; Score 30; DB 14; Length 265;

59.3%; Pred. No. 1.4;

Query Match 10.7%; Score 30; DB 22; Length 2088;
 Best Local Similarity 27.9%; Pred. No. 3.4;
 Matches 46; Conservative 36; Mismatches 83; Indels 0; Gaps 0;

QY 65 gcttgagtcagcaaacactttccacgctcaccaatccaggtgagtagctgtgagga 124
 ||||| : : : : : ||||| :
 Db 190 GYTNCCRTTYTCTARTGATYTCNACRTARTCSNWSNARNARNCCTRTTTNARDA 131
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 125 agaaagcgcgaagcctctatctaccgtccctctgtttgtttgtcgcgacgcgctgagtt 184
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 130 TNARRTCNARDATFAAYTTNSWNCRAANGCYTCDATYTGRAANSWNGCYTNGCNARRT 71
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 185 gatcggcgcaaaacctctgtaagtcocgctgctaagataaa 229
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 70 GNACNGCYTTRTTRTGTYTGTGTGRTGCKNGCYTTNGTRTCA 26
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

RESULT 14
 AAT11104
 ID AAT11104 standard; cDNA; 2272 BP.

XX AC AAT11104;

XX DT 09-AUG-1996 (first entry)

XX DE Transforming growth factor beta MP-121 cDNA.

XX KW TGF-beta; MP-121; mitogen; differentiation; induction; promotion;
 KW maintenance; morphogen; tissue regeneration; dental implantation;
 KW wound healing; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 128..1186

XX FT /*tag= a

XX FT /product= TGF-beta_MP-121

XX FT 836..1183

XX FT /*tag= b

XX DE19511243-A1.

XX PD 04-JAN-1996.

XX PF 27-MAR-1995; 95DE-1011243.

XX PR 01-JUL-1994; 94DE-1423190.

XX PA (BIOP-) BIOPHARM GES BIOTECHNOLOGISCHEN ENTWICKL.

XX PI Bechtold R, Neidhardt H, Pohl J, Hoetten G;

XX WPI; 1996-050788/06.

XX P-PSDB; AAR89729.

XX PT DNA encoding transforming growth factor beta MP-121 - has mitogenic
 XX and differentiation-inducing activity, e.g. for use in wound healing

XX PS Claim 1; Page 10; 15pp; German.

XX CC A cDNA library prepared using total RNA from human liver was
 CC subjected to PCR amplification using primers corresp. to conserved
 CC regions within the TGF-beta family. Amplification products were
 CC subcloned and sequenced; one clone (designated PSK-MP121) was found
 CC to contain a new sequence. Part of the insert from this clone was
 CC used to re-screen the human liver cDNA library and a 2272 bp fragment
 CC (i.e. the present sequence) coding for a TGF-beta-like protein was
 CC isolated. The protein encoded by the cDNA insert has mitogenic and
 CC differentiation-inducing properties making it (or fusion proteins
 CC comprising it or heterodimers of the protein with a cysteine knot
 CC motif protein) useful for inducing tissue regeneration.

XX SQ Sequence 2272 BP; 510 A; 663 C; 513 G; 586 T; 0 other;

Query Match 10.7%; Score 30; DB 17; Length 2272;
 Best Local Similarity 59.3%; Pred. No. 3.5;
 Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 192 qcqcaaaacactctctgtaagtcocgctgctaaagataaaatagcgcgggctagtattac 251
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 940 qcacatgaactctctgcatagggcagtcgcccactacacatagcagcgtgctgtattgc 999
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 252 ccccgcggtccccaacacctctgctaa 277
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1000 tgctccctttcacactgcagtcctca 1025
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15

AAF74421

ID AAF74421 standard; DNA; 2272 BP.

XX AC AAF74421;

XX DT 08-MAY-2001 (first entry)

XX DE Human TGF-beta MP121 nucleotide sequence SEQ ID NO:3.

XX KW Human; transforming growth factor-beta; monomeric protein; MP52;
 KW MP121; dimeric protein; TGF-beta; vulnerary; antiulcer; nontropic;
 KW neuroprotective; antiinfectivity; osteopathic; gene therapy; bone;
 KW cartilage; dental; wound healing; connective tissue; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 128..1186

XX FT /*tag= a

XX FT /product= "MP121"

XX FT /note= "transforming growth factor (TGF)-beta monomeric
 XX protein"

XX DE1074620-A1.

XX PD 07-FEB-2001.

XX PF 06-AUG-1999; 99EP-0115613.

XX PR 06-AUG-1999; 99EP-0115613.

XX PA (HYGE-) HYGENE AG C/O MAEDER & BAUMGARTNER TREUH.

XX WPI; 2001-228100/24.

XX P-PSDB; AAB70530.

XX PT Novel monomeric protein of transforming growth factor-beta family for
 XX prevention or therapy of diseases associated with bone, cartilage
 XX damage, promotion of wound healing, has substitution or deletion of
 XX cysteine -

XX PS Claim 16; Page 17-20; 31pp; English.

XX CC The present invention describes a protein (I) selected from the members
 CC of the transforming growth factor-beta (TGF-beta) superfamily, which is
 CC monomeric due to substitution or deletion of a cysteine which is
 CC responsible for dimer formation. Also described are: (1) nucleic acid
 CC (II) encoding (I); (2) expression vector (III) containing (II) in a
 CC suitable vector system; (3) host cell (IV) containing (III) capable of
 CC producing (I); and (4) a pharmaceutical composition (V) containing (I),
 CC (II), (III) or (IV). (I) has vulnerary, antiulcer, nontropic,
 CC neuroprotective, antiinfectivity and osteopathic activities, and can be
 CC used in gene therapy. (V) is useful for the prevention or therapy of
 CC diseases for which also the dimeric form of the protein would be
 CC indicated. Diseases treatable include diseases associated with bone
 CC and/or cartilage damage or affecting bone and/or cartilage disease or
 CC situations in which cartilage and/or bone growth is desirable, for spinal

CC fusion, for damaged or diseased tissue associated with connective tissue
 CC including tendon and/or ligament, periodontal or dental tissue including
 CC dental implants, neural tissue including CNS tissue and neuropathological
 CC situations, tissue of the sensory system, liver, pancreas, cardiac,
 CC blood vessel, renal, uterine and thyroid tissue, skin, mucous membrane,
 CC endothelium, epithelium, for promotion or induction of nerve growth,
 CC tissue regeneration, angiogenesis, wound healing including ulcers, burns,
 CC injuries or skin grafts, induction of proliferation of progenitor cells
 CC or bone marrow cells, for maintenance of a state of proliferation or
 CC differentiation, for treatment or preservation of tissue or cells for
 CC organ or tissue transplantation, for integrity of gastrointestinal lining
 CC and for treatment of disturbances in fertility, contraception or
 CC pregnancy. The present sequence encodes the specifically claimed
 CC TGF-beta monomeric protein MP121, from the present invention.
 XX
 SQ Sequence 2272 BP; 510 A; 663 C; 511 G; 585 T; 3 other;

Query Match 10.7%; Score 30; DB 22; Length 2272;
 Best Local Similarity 59.3%; Pred. No. 3.5;
 Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
 QY 192 cgccaaaacacctgtgcttaagtcctgcttaagataaaatagccgggggctagtattac 251
 ||||| ||||| || ||||| || ||||| || ||||| || ||||| ||
 Db 940 cgccatgaacttctgcatagggcagtcgcccactacacatagcagcagcatgctgtattgc 999
 QY 252 ccccgcggtcccccacacctctgctaa 277
 || | || | || || |
 Db 1000 tgccctcttccacactgcagtgctca 1025

Search completed: June 19, 2002, 17:14:11
 Job time: 18909 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 16:22:15 ; Search time 247.21 Seconds
(without alignments)
279.208 Million cell updates/sec

Title: US-09-462-955B-1_COPY_711_991

Perfect score: 281

Sequence: 1 aatcgagttatggcgggcc.....ccccaaacctctgtaacccc 281

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	191	68.0	1291	4	US-09-462-975-1
2	32.2	11.5	4411529	4	US-09-103-840A-1
3	31	11.0	31	4	US-09-462-975-2
4	30	10.7	265	3	US-08-289-222E-5
5	30	10.7	265	4	US-09-054-526B-5
6	30	10.7	2088	4	US-09-351-414-3
7	30	10.7	2272	1	US-08-482-577B-1
8	30	10.7	2272	3	US-08-289-222E-2
9	30	10.7	2272	4	US-09-218-176-1
10	30	10.7	2272	4	US-09-054-526B-2
11	28.2	10.0	1414	4	US-08-686-968C-228
12	27.4	9.8	2202	1	US-08-655-836-2
13	27.4	9.8	2202	2	US-09-020-753-2
14	27.4	9.8	2364	1	US-08-655-836-3
15	27.4	9.8	2364	2	US-09-020-753-3
16	27.4	9.8	2799	3	US-08-968-752B-5
17	27.2	9.7	1381	4	US-09-426-557-5
18	27.2	9.7	1463	4	US-09-426-557-1
19	27.2	9.7	1478	4	US-09-426-557-7
20	27.2	9.7	1541	4	US-09-426-557-3
21	26.8	9.5	801	4	US-09-020-956-16
22	26.8	9.5	801	4	US-09-030-607-16
23	26.8	9.5	801	4	US-09-439-313-16
24	26.8	9.5	3526	1	US-08-323-084A-7
25	26.8	9.5	3526	1	US-08-323-084A-8
26	26.8	9.5	3526	1	US-08-674-008-7
27	26.8	9.5	3526	1	US-08-674-008-8

28	26.6	9.5	447	4	US-09-202-712-19	Sequence 19, Appli
29	26.6	9.5	1322	1	US-08-323-531-1	Sequence 1, Appli
30	26.6	9.5	1322	1	US-08-198-094-1	Sequence 1, Appli
31	26.6	9.5	1322	4	US-08-107-794A-1	Sequence 1, Appli
32	26.6	9.5	1322	5	PCT-US93-07424-1	Sequence 1, Appli
33	26.6	9.5	1322	5	PCT-US95-02087-1	Sequence 6, Appli
34	26.6	9.5	1638	4	US-07-861-458C-6	Sequence 5, Appli
35	26.6	9.5	4093	4	US-07-861-458C-5	Sequence 10, Appli
36	26.4	9.4	1111	1	US-08-202-186-10	Sequence 10, Appli
37	26.4	9.4	1237	4	US-09-007-678B-48	Sequence 48, Appli
38	26.4	9.4	3129	4	US-09-387-695-1	Sequence 1, Appli
39	26.4	9.4	4245	5	PCT-US96-05320A-516	Sequence 516, App
40	26.4	9.4	9717	4	US-09-251-645-1	Sequence 1, Appli
41	26.4	9.4	17327	1	US-07-906-871-15	Sequence 15, Appli
42	26.2	9.3	740	4	US-09-020-956-17	Sequence 17, Appli
43	26.2	9.3	740	4	US-09-030-607-17	Sequence 17, Appli
44	26.2	9.3	740	4	US-09-439-313-17	Sequence 17, Appli
45	26.2	9.3	751	4	US-09-020-956-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1
US-09-462-975-1
; Sequence 1, Application US/09462975
; Patent No. 6303345
; GENERAL INFORMATION:
; APPLICANT: Rohde, Wolfgang
; APPLICANT: Becker, Dieter
; APPLICANT: Randles, John W.
; APPLICANT: Hehn, Alain
; APPLICANT: Salamini, Francesco
; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER
; FILE REFERENCE: 23232.000301
; CURRENT APPLICATION NUMBER: US/09/462,975
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/EP98/04345
; PRIOR FILING DATE: 1998-07-13
; PRIOR APPLICATION NUMBER: 19730502.4
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1291
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:/note=synthetic
; OTHER INFORMATION: construct
US-09-462-975-1

Query Match	68.0%	Score 191;	DB 4;	Length 1291;
Best Local Similarity	100.0%;	Pred. No. 4.1e-63;		
Matches 191;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	aatcgagttatggcgggccacaaataaaagattccatttgataagaacgaatctgtta	60	
Db	1101	aatcgagttatggcgggccacaaataaaagattccatttgataagaacgaatctgtta	1160	
QY	61	ctttgcttgagtcgacgcaaccactttccacgtccacaaatccaggtagagcttgcg	120	
Db	1161	ctttgcttgagtcgacgcaaccactttccacgtccacaaatccaggtagagcttgcg	1220	
QY	121	gagaagaagccgaagcctctatctaccgtccattgtttttgctgagcgagctg	180	
Db	1221	gagaagaagccgaagcctctatctaccgtccattgtttttgctgagcgagctg	1280	
QY	181	agttgactcg	191	
Db	1281	agttgactcg	1291	

```

Db      31  AGCCGCGGGGCTAGTATTACCCCGGGCT 1
|||||
RESULT 4
US-08-289-222E-5
; Sequence 5, Application US/08289222E
; Patent No. 6120760
; GENERAL INFORMATION:
; APPLICANT: HOTTEN, GERTRUD
; APPLICANT: NEIDHARDT, HELGE
; APPLICANT: BECHTOLD, ROLF
; APPLICANT: POHL, JENS
; TITLE OF INVENTION: GROWTH/DIFFERENTIATION FACTORS OF THE TGF-B
; TITLE OF INVENTION: FAMILY
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAIIDO, MARMEI STEIN, MURRAY & ORAM
; STREET: 655 FIFTEENTH STREET, N. W., G STREET LOBBY,
; SUITE: SUITE 330
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,222E
; FILING DATE: 25-AUG-1999
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,222
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 23 190.3
; FILING DATE: 07-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EPO 92102324.8
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/00350
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KITTS, MONICA CHIN
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P564-9021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA from mRNA
US-08-289-222E-5

```

	Query Match	10.7%	Score 30;	DB 3;	Length 265;
	Best Local Similarity	59.3%;	Pred. No. 0.066;		
	Matches 51;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;
Qy	192	cgccaaacactctgctaagtcocgtgctaagtataaataggccggggcgtatattac	251		
Dd	19	CGCCATCAACTCTGTCATAGGGCAGTGCCTACTACACATACAGGCATGCCCTGGTATTGC	78		
Qy	252	cccgcgggctcccccaacctctgctaa	277		
Dd	79	TGCTCCTCTTTACACACTGCAGTGCTCA	104		

1

; PRIOR APPLICATION DATA:

Qy	4	cgagttatggcggggcccaataaaagagattccaattggataaagaacgaactgttactt	63
Db	737		
Qy	64	tgcttgcagtgacgcgaacaccatttccac	92
Db	797		825
Qy	796	caagcaatggggcaagcctctgtacaaaatatcaaaatgggaagaaggaattattaagt	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 14:04:28 ; Search time 7489.97 Seconds
(without alignments)
506.363 Million cell updates/sec

Title: US-09-462-955b-1_COPY_711_991
Perfect score: 281
Sequence: 1 aatcagattatggcgccgccc.....ccccaaacctcgttaacccc 281

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthm: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_gss: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vit: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34.4	12.2	898	9	AL667528
2	34	12.1	1101	12	CNS000BES
3	33.2	11.8	1030	9	BB610652
4	33	11.7	644	12	AZ339575
5	33	11.7	762	9	AV763886
6	33	11.7	977	12	A0459250
7	32.8	11.7	668	12	BH437599
8	32.6	11.6	659	9	AA221432
9	32	11.4	1037	12	CNS0374V
10	32	11.4	1210	11	AK010602
11	31.8	11.3	533	9	AW450316
12	31.8	11.3	578	10	C79787
13	31.6	11.2	335	10	BE246832
14	31.6	11.2	358	9	AI205836
15	31.6	11.2	410	9	AA834925
16	31.4	11.2	416	9	AI016082
17	31.4	11.2	424	9	AA889602

C 18	31.2	11.1	297	9	BB295841
C 19	31.2	11.1	299	9	BB285126
C 20	31.2	11.1	391	9	AW037599
C 21	31.2	11.1	617	12	BH087127
C 22	31	11.0	298	9	AW014292
C 23	31	11.0	620	12	BH540957
C 24	31	11.0	625	12	AQ874985
C 25	31	11.0	692	12	AG145182
C 26	31	11.0	694	12	BH451673
C 27	31	11.0	768	9	AL666434
C 28	30.8	11.0	440	9	AA438686
C 29	30.8	11.0	533	10	BH198148
C 30	30.8	11.0	713	10	BT552821
C 31	30.8	11.0	875	10	BE962362
C 32	30.8	11.0	918	10	BE378992
C 33	30.8	11.0	1101	12	CNS001C3
C 34	30.6	10.9	228	9	AV308279
C 35	30.6	10.9	319	9	AV104431
C 36	30.4	10.8	247	12	AQ025371
C 37	30.4	10.8	260	9	AI336966
C 38	30.4	10.8	260	9	AI337129
C 39	30.4	10.8	260	9	AI624907
C 40	30.4	10.8	289	9	AI825120
C 41	30.4	10.8	340	9	AA051125
C 42	30.4	10.8	373	12	AZ812694
C 43	30.4	10.8	389	12	AZ510706
C 44	30.4	10.8	633	9	AV689499
C 45	30.4	10.8	795	12	AQ876011

ALIGNMENTS

RESULT 1

AL667528
LOCUS AL667528 898 bp mRNA linear EST 11-JAN-2002
DEFINITION clone 020ZG05 5', mRNA sequence.
ACCESSION AL667528.1 GI:18134435
VERSION EST.
KEYWORDS Ciona intestinalis.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis.
REFERENCE 1 (bases 1 to 898)
AUTHORS Genoscope.
TITLE Ciona intestinalis directional larval cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
IMPORTANT: this sequence may contain errors. The Ciona intestinalis library from which the clone was isolated may be contaminated with cDNAs from bacteria or other Eukarya.
Directional larval cDNA library originate from Dr M.Branno, Stazione A.Dohrn, Naples, Italy, and was prepared in pBluescript2SK+.

FEATURES
source 1, 898
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="020ZG05"
/clone.lib="directional larval cDNA library"
/note="Vector: pBluescript2SK+" 1 others
BASE COUNT 289 a 186 c 157 g 265 t
ORIGIN

Query Match 12.2%; Score 34.4; DB 9; Length 898;
Best Local Similarity 60.9%; Pred. No. 1.5;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

<p>electrophoresis. Vector DNA was prepared from a derivative of pMD42 (G1/4732114 gb AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent <i>E. coli</i> XL10-Gold (Stratagene) cells and selected for ampicillin resistance.</p>										
BASE COUNT	177 a	186 c	100 g	181 t						
ORIGIN										
Query Match	11.7%	Score 33	DB 12	Length 644						
Best Local Similarity	60.7%	Pred. No. 4.1								
Matches	54	Conservative	0	Mismatches	35	Indels	0	Gaps	0	
QY	58	ttactttgcttcagtcgacgcacacactttccacgtccacatccacgttgagctctg	117							
Db	354	TTTCTGTGCACTGAGTGTAGGTTATCTCTCTCTCTTACAAGCCTAAGGCATTCTCTGG	295							
QY	118	ctggagagaagaagccgcaagcctctctatc	146							
Db	294	CTTTAATAAAGAGCTGCAAGCCTATATCT	266							
RESULT	5									
AV763886										
LOCUS	AV763886	762 bp	mRNA	linear	EST 19-OCT-2000					
DEFINITION	AV763886	MDS Homo sapiens cDNA clone	MDSEE03 5'	mRNA sequence.						
ACCESSION	AV763886									
VERSION	AV763886.1	GI:10921734								
KEYWORDS	EST.									
SOURCE	human.									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.									
AUTHORS	Gu, J., Zhao, M., Huang, Q., Xu, X., Li, Y., Peng, Y., Song, H., Xiao, H., Gu, Y., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang, Y., Gao, G., Zhang, Q., Chen, S., Han, Z. and Chen, Z.									
TITLE	Homo sapiens cDNA MDS clones									
JOURNAL	Unpublished (2000)									
COMMENT	Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai.									
FEATURES	Location/Qualifiers									
source	1..762									
	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
	/clone="MDSEE03"									
	/clone_lib="MDS"									
	/tissue_type="Bone marrow"									
	/cell_type="CD34+ hematopoietic stem/progenitor cell"									
	/lab_host="BM25.8"									
	/note="Vector: pTriplex2; Site_1: sfiIA; Site_2: sfiIB"									
	/notes="Vector: pTriplex2; Site_1: sfiIA; Site_2: sfiIB"									
BASE COUNT	203 a	172 c	177 g	209 t	1 others					
ORIGIN										
Query Match	11.7%	Score 33	DB 9	Length 762						
Best Local Similarity	48.9%	Pred. No. 4.3								
Matches	87	Conservative	0	Mismatches	91	Indels	0	Gaps	0	
QY	36	catttggatagaagacgaatctgttactttgcttcagtcgacgaacacatttcacgctc	95							
Db	471	CAGCTTCGCGAAGTTACCTGATTGTATTATTAGAATGATACCACTCTGCTGATTC	530							

Qy 98 caatccaggtgagtagcttgctggagaaagccgcaagcctctatctaccgtccattt 157
 Db 61 CACACAACCTTCAGTAGGTTGATGGCTTCTACTGGTCCAGGCCCTTGACCCCGAGCTAGGG 120
 Qy 158 gttttgcgcgatcggaacggt 179
 Db 121 CCTCAGGGACCTTCGGGCGTT 142

Search completed: June 19, 2002, 14:04:31
 Job time: 7529 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 15:49:33 ; Search time 5904.86 Seconds
(without alignments)
184.285 Million cell updates/sec

Title: US-09-462-955B-1_COPY_734_785

Perfect score: 52

Sequence: 1 aataaagattccatttggg.....tgttactttgttgatgctgc 52

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_ov.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vt.*

15: gb_vt.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_mu.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vt.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
------------	-------------	--------------	-------	-------------

ALIGNMENTS

RESULT 1

ARI72770

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

linear PAT 17-DEC-2001

DNA

1291 bp

Sequence 1 from patent US 6303345.

ARI72770

ARI72770.1 GI:17912261

Unknown.

Unknown.

Unclassified.

1 (bases 1 to 1291)

Rohde,W., Becker,D., Randles,J.W., Hehn,A. and Salamini,F.

Use of a virus DNA as promoter

Patent: US 6303345 A 1 16-OCR-2001;

Location/Qualifiers

1..1291

/organism="unknown"

336 a 323 c 332 g 300 t

Query Match 100.0%; Score 52; DB 6; Length 1291;
Best Local Similarity 100.0%; Pred. No. 5.6e-07;

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aataaagattccattgataagaacgaatctgttaacttgcagtc 52
|||||
Db 1124 AATAAAGATTCATTGGATAAGAACGAATCTGTTACTTGTTCGAGTGC 1175

RESULT 2

CFDCG Coconut foliar decay virus, complete genome.
LOCUS 1291 bp ss-DNA circular VRL 02-AUG-1993

DEFINITION Coconut foliar decay virus, complete genome.
ACCESSION M29963
VERSION M29963.1 GI:323306

KEYWORDS circular; complete genome.
SOURCE Coconut foliar decay virus DNA, clones NMORG, PORG, and PORGSUA3A.
ORGANISM Coconut foliar decay virus

REFERENCE 1 (bases 1 to 1291)
1 (bases 1 to 1291)
Rohde, W., Randles, J.W., Langridge, P. and Hanold, D.

AUTHORS Nucleotide sequence of a circular single-stranded DNA associated
TITLE with coconut foliar decay virus

JOURNAL Virology 176, 648-651 (1990)
MEDLINE 90266484
COMMENT Draft entry and printed sequence for [1] kindly submitted by
W. Rohde, 15-MAR-1989, for release after publication.

FEATURES
Source
1..1291
/organism="Coconut foliar decay virus"
/db_xref="taxon:12474"

stem_loop
40..70
/note="stem-loop structure"

CDS
103..975
/note="ORF 1"
/codon_start=1
/protein_id="AAA42894.1"

translation="MGSSIRWCFTLNYETEEAANVRRRIESLNLYAIVGDEVAPS
TRGKRLQFIHLKTRGRRLQGLVLDNRIHLEPTRGDEQNRDVCSEKRVLLHGV
TPGVKRFPAQAEDELRLIEDPGYRCVHVHCASVETRWAAENPFPPYHNWQ
LEVLSAIGEPADRTILWTCGRDGGKSVAKYLGKLPDWFYTCGGTRKDVLYQIE
DKRNLIDVPCRNLELYNALLECVKRNATSSDKYEPLSYLGFDPVHVLVFNVLDP
YKISDRIRIKLWNI"

314..775
/note="ORF 2"
/codon_start=1
/protein_id="AAA42895.1"

translation="MTGFTWSPVPTNRIETTVRNGCFSTESRLVLESKGDWPN
DLRLNLSAWKTDATEDALYELRWNGQDGLKIRSHFHTITGLKCCLRSESQRT
IAQSGYADETEETGSPCLPNISDSPTGTHVPEPTYCTSTRTONEI"

complement(422..568)
/note="ORF 6"
/codon_start=1
/protein_id="AAA42896.1"

translation="MEMGTDFQRPILSPKLRVORIFGIRLPGGVHVPQQIVGP
IVAF"

639..797
/note="ORF 3"
/codon_start=1
/protein_id="AAA42897.1"

translation="MTRRRRRREVRCQISRTQARLVLMHWNQGRIVPVHRGPKTK
FNPRCTQV"

complement(823..987)
/note="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"

translation="MPTHNLNIPQFVSVPADFIIRODIGKYEYMHMVPEKIKTKGFVFF
RTECPVLNTF"

1098..1286
/note="ORF 4"

1098..1286
/note="ORF 4"

1098..1286
/note="ORF 4"

1098..1286
/note="ORF 4"

1098..1286
/note="ORF 4"

/codon_start=1
/protein_id="AAA42899.1"

/db_xref="GI:323312"
/translation="MNRVMGGPTIKDSIWIRTNLLCLQCTQPLSTSPIOVSSLEKK
AASLYLPSICFCAIGRLS"

BASE COUNT 336 a 323 c 332 g 300 t
ORIGIN

Query Match 100.0%; Score 52; DB 14; Length 1291;
Best Local Similarity 100.0%; Pred. No. 5,6e-07;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aataaagattccattgataagaacgaatctgttaacttgcagtc 52
|||||
Db 1124 AATAAAGATTCATTGGATAAGAACGAATCTGTTACTTGTTCGAGTGC 1175

RESULT 3

AP004548/c

LOCUS

DEFINITION

AP004548

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Lotus japonicus

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE


```

KEYWORDS      synthetic construct.
SOURCE        synthetic construct.
ORGANISM      artificial sequence.
REFERENCE     1 (bases 1 to 243)
AUTHORS      Rance, I., Theisen, M. and Gruber, V.
TITLE        Chimeric expression promoters originating from commelina yellow
             mottle virus and cassava vein mosaic virus
JOURNAL      Patent: WO 0058485-A 1 05-OCT-2000;
             MERISTEM THERAPEUTICS (FR); RANCE IANN (FR); THEISEN MANFRED (FR)
FEATURES     Location/Qualifiers
             1..243
             /organism="synthetic construct"
             /db_xref="taxon:32630"
             /note="243 bp promoter fragment from the intergenic region
             of Commelina Yellow Mottle Virus"
             1..243
             promoter          71 a 53 c 45 g 74 t
BASE COUNT   71 a 53 c 45 g 74 t
ORIGIN
Query Match      50.8%; Score 26.4; DB 6; Length 243;
Best Local Similarity 69.2%; Pred. No. 1.2e+02;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 aataaaagattccattggataaagacgaatctgttacttctgtcagtcg 52
   ||| ||||| ||| ||||| ||||| ||| ||| ||||| |||||
Db 154 AAGATAGATTGCTTTTGATAGACGACACAACTTATGCCCTTCAGTCG 103

RESULT 5
CYMVCG/c      CYMVCG      7489 bp DNA linear VRL 12-SEP-1993
LOCUS         Commelina yellow mottle virus complete circular genome.
DEFINITION   X52938
VERSION      X52938.1 GI:59047
KEYWORDS     complete genome; protease; reverse transcriptase; ribonuclease H.
SOURCE       Commelina yellow mottle virus.
ORGANISM     Commelina yellow mottle virus.
             Viruses; Retroid viruses; Badnavirus.
REFERENCE    1 (bases 1 to 7489)
AUTHORS      Olszewski, N.E.
TITLE        Direct Submission
JOURNAL      Submitted (02-WAY-1990) Olszewski N.E., University of Minnesota,
             Dept of Plant Biology, 220 Biological Sciences Center, 1445 Gortner
             Ave., St. Paul, MN 55108, USA
             (revised by [4])
REMARK       2 (bases 1 to 7489)
AUTHORS      Medberry, S.L., Lockhart, B.E. and Olszewski, N.E.
TITLE        Properties of Commelina yellow mottle virus's complete DNA
             sequence, genomic discontinuities and transcript suggest that it is
             a pararetrovirus
JOURNAL     Nucleic Acids Res. 18 (18), 5505-5513 (1990)
MEDLINE     91016835
REFERENCE    3 (bases 1 to 7489)
AUTHORS      Olszewski, N.E.
TITLE        Direct Submission
JOURNAL     Submitted (29-NOV-1990) Olszewski N.E., University of Minnesota,
             Dept of Plant Biology, 220 Biological Sciences Center, 1445 Gortner
             Ave., St. Paul, MN 55108, USA
             4 (bases 1 to 7489)
AUTHORS      Olszewski, N.
TITLE        Direct Submission
JOURNAL     Submitted (07-DEC-1990)
COMMENT     [3] The extra c at position 2672 permits readthrough to result in
             one ORF (216 kd protein) rather than two (47 kd ORF and 163 kd ORF).
FEATURES     Location/Qualifiers
             1..7489
             /organism="Commelina yellow mottle virus"
             /db_xref="taxon:10653"
             /clone="pCoYMV89 (and pCoYMV89)"

```

```

1..23
/note="put. tRNA-Met (put. primer for minus strand
synthesis)"
/product="tRNA-Met"
300..302
/note="gac was cga"
/citation=[1]
496..1098
/note="ORF 1"
/codon_start=1
/protein_id="CAA37108.1"
/db_xref="GI:59048"
/db_xref="SWISS-PROT:P19200"
/translation="MNWLKSHPTPLGLLPYSLDPFCFMNQVDQVKQLIDMLSSA
KKLSEVIVFTPEVKINRLDLAHNIHIAHRVALGFKVYLYLVLDIIPPLKNIQKSO
KESSNLQSVLKIVKEORSLKQIEDQLSKVQSELAKREDYLSRRLSKQDVELVV
RISEQPKIEKQTEALTEELKLVKEEVAKLIHSPKGMVLN"
541
/note="t in clone pCoYMV100"
873
/note="t in clone pCoYMV100"
1098..1505
/note="ORF 2"
/codon_start=1
/protein_id="CAA37109.1"
/db_xref="GI:59049"
/db_xref="SWISS-PROT:P19201"
/translation="MSNLTEKSGYKEALSVTNNYLAPAIIGGATDVLGTGVTKOLNT
TIYLVAKITQIEDLQSTIKLEERVQSLKAKTPVVTQDPNPEILSKLSDIOISLAR
QRAVPAISGVSNYSTPIKIKVDRILRVFKFN"
1338..1340
/note="3bp deletion in pCoYMV100"
1506..7166
/note="ORF 3, 216 kd protein"
/codon_start=1
/protein_id="CAA37110.1"
/db_xref="GI:59050"
/db_xref="SWISS-PROT:P19199"
/translation="MATRLPAVTQTDGTSRTATESGVPPEYEDQIRSYRNDQRRRIHWA
GRGRLLSLMPGVSSSERTLEOMNPEVQLOSRMNRARPAEVLVYRTFHGKFASTM
YSHRSEEMVMVNGSQVDSFTQESSEFEVLSTGTEFHIGVMVRIOLILKPKFAGTM
ALIVERTWSDDRAVLAAMEIDLSEGNQIVYVLPDLMWTIKSFYHFIOLCVMPKGYD
GWQEDNLLITRLGLTNSNVGFAYDKAWVEHLOSNGVKKAKGKBAKREKPHNG
QNIPEKVVYPMQPTKMAVSNVDTGTSLSRFSNYAAASTKPPYNEKKEDEINDEQ
EINSLNLIIDESTDEEYQYQYAWSQVDSFTYDTDGWEEIDRCNDLPEY
VPSETSTPIDSEALIDFELHAYEQRCDSQESQSDGPKRYEYTPQSSPEHLNDE
SPRSASSSTSMQDDVEIVELMKEMRMKOKKKAKOQALSSQAEKPIITEEN
KQAEPTQEEIPTHKEQPEIQNEEIHVEEPAFKHLAQLSELVMAESSGSG
VGQPPVNAQPDVNMGEFAPATSQATWSNGVNPVKSANFRWKPVGNPQLPSAQ
KDGAMLVFGMNYSPVEFDRWASITRNYISSFNFDGDKTAMMEDLLIGETKRLFVS
WMRFKDEYNLIKIANODGGTQAILSQIRIFLGEDPVLGQNTVONIAFRKLKOLVC
PNYQSTRYLYMDYLTAAETGLMWSETGPAISEELTKMPAAIGERVAQYKIMDPT
SAVNLPSRVYFTINVLTEOCKEASVMSRLKALDFCRDPTIEGYGRSGEKKYATKA
TKYTKAHDNHIRVTKAYQKCKCYICGQEGHYANQCNKHKDOORVAILOSILKE
NEEVSADKEDDEDEIFSGEEDYQETINVEEDDIIQIILKEFKGDLSSRWVG
PNFPGPAEYQMGVLKPSWRPPIQATLEEINCHNNTAISTQIACSCQKLAGVQ
CHHCHAVYCFMCAEAYHDQAEKILSKDYSFSGKKGKAVIEDEIEGEFLISQLQ
QBNRLOKQOVERLOELMKLHREKDEALKHSEKASRVEFTIOESDEAELNLIKELRQ
FKETPMATAOLKEALIVQEDDTIEERCAMILLEKHTENIYATKAEVNGLYNVKVG
IKPDNMEPYINAVTDGTACGLIQISALIPENYEDAKVTYVNFERSVLGIGSTOMIKA
GRILIGEQTFRMPTVYVMNGLSPQIMIGCSFIRSGEGGLRERDKDIITFKLVTSI
ETSRTQVANSTIELESEDEYLNIAASVETPSDFQEFARKNKDILKEMKEMKVI
NPMFEWKNNKIKCLNIPDIIKIMGRPIKHVTPGDEEAMTRQINLLQMKVIRPSES
KHRSTAFIVRSGTEPDITGKKGKRMVFNKLLNENTESDYSLPGINTIISKVG
RSKIYKFDLKSQFWQVMAFEESVPKTAFLACNKLVEMLVPEFLKNAPAFORKMDN
VFKGEKFLAVTIDILLVSETAHQSHLYTMLOCKENGLILSPTKMKIGTPIIDF
LGASGCTKIKLOPHIISKICDFDEKATLPEGMRSWGLSIFARNYQDIOGKLVQPL
ROKMAPTGDKRMNPETKVMVQIKKVKNLPLDQLPKDQSTIITGCGWGWAVCK
WMSKHDPRTICAYAGSENFPIKSTIDAEITQAAIHGLDKFKIYLDKELLIRSD
CEAIIKFYNTKNTENKPSRVRLTFSDFLTGLGITVTFEHDHCKHGLADLSRMFINI
VEKNDESPRFTSSVEDALKVCNDDHGRNLISAVINDIITVLR"
1970
/note="a in clone pCoYMV100"

```

```

variation      2452
/Note="t in clone pCoVMV100"
old_sequence   2672..2674
/Note="ccc was cc in [1]"
/citation=[1]
variation      2977
/Note="c in clone pCoVMV100"
variation      4611
/Note="a in clone pCoVMV100"
variation      4794
/Note="t in clone pCoVMV100"
variation      6302
/Note="t in clone pCoVMV100"
variation      6434
/Note="c in clone pCoVMV100"
variation      7378
/Note="c in clone pCoVMV100"
variation      7462
/Note="additional t in pCoVMV100"
variation      7478
/Note="c in clone pCoVMV100"
variation      7482
/Note="g in clone pCoVMV100"
BASE COUNT    2662 a 1389 c 1577 g 1861 t
ORIGIN

Query Match      50.8%; Score 26.4; DB 14; Length 7489;
Best Local Similarity 69.2%; Pred. No. 86;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 aataaagattcattgataagaacgaatctgttacttctgttcgttcgactgc 52
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7277 AAGATAAGATTCTTTTGTATGAAGACGAACAACCTTTATGCTTTCAGTGC 7226

RESULT 6
LOCUS      CEC25A1          40597 bp      DNA      linear      INV 24-JAN-2002
DEFINITION Caenorhabditis elegans cosmid C25A1, complete sequence.
ACCESSION  Z81038
VERSION     Z81038.1 GI:1627591
KEYWORDS    HTG; Aryl hydrocarbon receptor nuclear translocator; Isoleucyl-tRNA
            synthetase; Lectin; NADH-ubiquinone oxidoreductase B8 subunit;
            Nuclear factor 5 like; RNA binding protein; Serine/threonine
            protein kinase; Yeast ABD1 protein like; Yeast hypothetical 44.9 KD
            protein like.
SOURCE      Caenorhabditis elegans.
ORGANISM    Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (sites)
AUTHORS     none.
TITLE       Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
JOURNAL     Science 282 (5396), 2012-2018 (1998)
MEDLINE     99069613
REMARK      The C. elegans Sequencing Consortium.
REFERENCE   2 (bases 1 to 40597)
AUTHORS     Mortimore,B.J.
TITLE       Direct Submission
JOURNAL     Submitted (21-OCT-1996) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            jess@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.

```

This sequence is the entire insert of clone C25A1. The true right end of clone F16C3 is at 23255 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence Z81065. The end of this sequence (40494..40597) overlaps with the start of sequence AL032856.

For a graphical representation of this sequence and its analysis see:- <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=C25A1>

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

Location/Qualifiers

1..40597

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="I"

/clone="C25A1"

complement(join(893..1316,1650..1829,1913..2091,2159..2268,2316..2441,2547..2677,4012..4070))

/gene="C25A1.3"

complement(join(893..1316,1650..1829,1913..2091,2159..2268,2316..2441,2547..2677,4012..4070))

/gene="C25A1.3"

/note="predicted using Genefinder"

Similarity to Yeast ABD1 protein (SW:P32783)

cDNA EST EMBL:T01105 comes from this gene"

/codon_start=1

/protein_id="CAB02758.1"

/db_xref="GI:3874433"

/db_xref="SPTREMBL:Q9XVS1"

/translation="MMKEVLDAIPKSGAEAGFGHNKMSSEVASHYKNVLOVIEGRKESRIFFEMNNMNVKSQLINDAKQVNDGVNPNRYLDLACGKGDLKKWDLAGAKDVVMADVDVSIQAEERYKQMGYKKNITVQFIVADCTKENLEDRIENKDPDLVSCQFALHVSFVDEASARIPKNAVGNLAPGGVFITGLDADRIVSMRNGENGQFANFVCKITENVEELAEKGKVLFGAKFHSDQVNCDFEFLAYFLPKHLLLELDMLFFVHMFAPAIKRWLEPGRRLLETGLTPEKLSKSDDEYLEAKAKLDAFPEDERIKTMGTLSKSWEAICMVLVFGFRKKSEAEKTEEPATTPVAESESEKVEVTESEKEDEQDCBHQEAQTN"

join(3183..3279,3326..3424,3471..3726,3772..3979)

/gene="C25A1.1"

join(3183..3279,3326..3424,3471..3726,3772..3979)

/note="predicted using Genefinder"

cDNA EST Yk488a2.5 comes from this gene

cDNA EST Yk590d6.5 comes from this gene

cDNA EST Yk488a2.3 comes from this gene"

/codon_start=1

/protein_id="CAB02762.1"

/db_xref="GI:3874437"

/db_xref="SPTREMBL:O17594"

/translation="MSSGFVSTSELDEKKARQAEWKIRKPTDALTVPPEVCNKTLEQLKNNKDAKOLEIDEAKLKNMVRGIDESVFLSELDSTKRVVVMKMKREOPLIKELAVTHLAAQNPSSSRILKPFSTKVLGPPKQAQALSTAIRKRSTSEKKQEDVSSKSVKPEPVIKQIGALQALCEYDPSSDESADSDDEPETLSLLQTSKSAAGGCC"

complement(join(4537..4791,5002..5130,5716..5920,5987..6090))

/gene="C25A1.2"

complement(join(4537..4791,5002..5130,5716..5920,5987..6090))

/note="predicted using Genefinder"

Similarity to Rat hepatocyte nuclear factor 5 (TR:Q63248), contains similarity to Pfam domain: PF00250 (Fork head domain), Score=170.0, E-value=1.3e-47, N=1"

/codon_start=1

/protein_id="CAB02761.1"

/db_xref="GI:3874436"

/db_xref="SPTREMBL:O17593"

/translation="MSHKYVTHKVVYICONGKTIICLHSSLHILPFTFOLRMEEILRFQLSSNSLADSPFMIPLSFDTSIMSPTCEQPKQHSYIGLIAAILSSPOKMWILA EYEWIWNPEYFRSGAGWRNSIRHNLSLNDCFVKGRAANGKHGHWAVHAFVACVDF

* by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source

```
1. .63682
  Location/Qualifiers
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="X"
    /map="q21.1-21.33"
    /clone="RP3-346A12"
    /clone_lib="RPCI-3"
    misc_feature
      1. .63682
        /note="assembly_fragment:00926"
BASE COUNT 20610 a 11146 c 11344 g 20582 t
ORIGIN
```

```
Query Match 50.8%; Score 26.4; DB 2; Length 63682;
Best Local Similarity 75.0%; Pred. No. 72;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
QY 4 aaagattccattgataagaacgaatctgttacttcttgc 47
||||| ||||| ||| | ||||| ||||| |||||
```

```
Db 58003 AAAATAATACATTGGTAGGTGCTAATCTGTACTGTATTGC 58046
```

RESULT 8

```
HS321115 75793 bp DNA linear PRI 23-NOV-1999
LOCUS Human DNA sequence from PAC 321115 on chromosome Xq21.3.
DEFINITION AL008632
ACCESSION AL008632.1 GI:3036776
VERSION Xq21.3.
KEYWORDS human.
SOURCE
```

ORGANISM

```
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 75793)
```

REFERENCE

AUTHORS

TITLE

JOURNAL

```
Direct Submission
Submitted (28-OCT-1997) Chromosome X Project Group
(http://www.sanger.ac.uk/HGP/ChrX/) Sanger Centre, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 8, 1998 this sequence version replaced gi:2578085.

```

COMMENT

IMPORTANT: This sequence is the entire insert of clone 321115. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variations annotated may not be found in the sequence submission corresponding to the overlapping clone as we submit sequences with only a small overlap as described above. This sequence was generated from part of the Sanger Centre chromosome X mapping group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX/>. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The true left end of clone 321115 is at 1 in this sequence. The true right end of clone 321115 is at 75793. 321115 is from the library RPC11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>.

FEATURES

source

```
Location/Qualifiers
  1. .75793
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="X"
    /map="Xq21.3"
    /clone="RP1-321115"
```

```
repeat_region
  1184. .1696
    /note="LIPAS repeat: matches 378. .892 of consensus"
repeat_region
  2499. .2800
    /note="AluSq repeat: matches 1. .302 of consensus"
repeat_region
  4436. .4469
    /note="17 copies of 2 mer 82 % conserved"
repeat_region
  5196. .5346
    /note="MER4A2 repeat: matches 1. .152 of consensus"
repeat_region
  5347. .5646
    /note="MER4A2 repeat: matches 211. .506 of consensus"
repeat_region
  6135. .6220
    /note="MER34 repeat: matches 530. .450 of consensus"
repeat_region
  7224. .7954
    /note="LTR8 repeat: matches 690. .1 of consensus"
repeat_region
  9239. .9272
    /note="17 copies of 2 mer 82 % conserved"
repeat_region
  9487. .9788
    /note="AluSq repeat: matches 1. .302 of consensus"
repeat_region
  10896. .11248
    /note="THE1C repeat: matches 370. .1 of consensus"
repeat_region
  11987. .12612
    /note="LTR8 repeat: matches 1. .558 of consensus"
repeat_region
  13703. .14143
    /note="MER4B repeat: matches 611. .194 of consensus"
repeat_region
  14353. .14597
    /note="MER4B repeat: matches 219. .1 of consensus"
repeat_region
  16357. .16402
    /note="23 copies of 2 mer 89 % conserved"
repeat_region
  16825. .16950
    /note="AluJo repeat: matches 136. .11 of consensus;
incomplete repeat"
repeat_region
  18670. .18972
    /note="AluJo repeat: matches 301. .2 of consensus"
repeat_region
  19098. .19184
    /note="AluJb repeat: matches 1. .87 of consensus;
incomplete repeat"
repeat_region
  19189. .19340
    /note="MER39 repeat: matches 311. .460 of consensus"
repeat_region
  19990. .20292
    /note="AluSq repeat: matches 1. .303 of consensus"
repeat_region
  21851. .21989
    /note="LIPB2 repeat: matches 183. .41 of consensus"
repeat_region
  21851. .21953
    /note="LIMD2 repeat: matches 183. .81 of consensus"
repeat_region
  22141. .22409
    /note="LIME3 repeat: matches 911. .635 of consensus"
repeat_region
  23025. .23532
    /note="2 copies of 254 mer 100 % conserved"
repeat_region
  23884. .24289
    /note="MTLD repeat: matches 57. .503 of consensus"
repeat_region
  26286. .26319
    /note="17 copies of 2 mer 88 % conserved"
repeat_region
  26910. .26955
    /note="23 copies of 2 mer 83 % conserved"
repeat_region
  28242. .28531
    /note="AluY repeat: matches 1. .290 of consensus"
repeat_region
  29316. .29645
    /note="AluJo repeat: matches 302. .1 of consensus"
repeat_region
  30847. .30979
    /note="FLAM_C repeat: matches 133. .1 of consensus"
repeat_region
  30997. .31030
    /note="17 copies of 2 mer 82 % conserved"
repeat_region
  32102. .32229
    /note="AluJo repeat: matches 129. .1 of consensus;
incomplete repeat"
repeat_region
  32751. .33112
    /note="MT1A2 repeat: matches 1. .374 of consensus"
repeat_region
  33120. .33192
    /note="MER5B repeat: matches 155. .84 of consensus"
repeat_region
  33812. .34808
    /note="L1 repeat: matches 4372. .5384 of consensus"
repeat_region
  34668. .35665
```

repeat_region /note="L1MA4 repeat: matches 4. .1046 of consensus"
36335. .36634
repeat_region /note="AluJo repeat: matches 1. .302 of consensus"
38132. .38448
repeat_region /note="MER4B repeat: matches 611. .297 of consensus"
38453. .38682
repeat_region /note="MER4B repeat: matches 231. .1 of consensus"
39063. .39247
repeat_region /note="AluJb repeat: matches 1. .177 of consensus;
incomplete repeat"
39702. .39837
repeat_region /note="AluJb repeat: matches 166. .301 of consensus;
incomplete repeat"
39884. .40234
repeat_region /note="MTJ1A1 repeat: matches 365. .1 of consensus"
40875. .41053
repeat_region /note="MER5B repeat: matches 1. .173 of consensus"
41288. .41418
repeat_region /note="AluJo repeat: matches 132. .1 of consensus;
incomplete repeat"
41628. .41914
repeat_region /note="L1PB3 repeat: matches 896. .591 of consensus"
41746. .42317
repeat_region /note="L1MA5 repeat: matches 758. .136 of consensus"
42058. .42427
repeat_region /note="L1PB2 repeat: matches 395. .29 of consensus"
42411. .44898
repeat_region /note="L1 repeat: matches 5377. .2924 of consensus"
44915. .45196
repeat_region /note="AluJo repeat: matches 17. .293 of consensus;
incomplete repeat"
45240. .45848
repeat_region /note="L1 repeat: matches 2875. .2257 of consensus"
45872. .46170
repeat_region /note="AluJb repeat: matches 5. .302 of consensus"
46258. .46557
repeat_region /note="AluJo repeat: matches 300. .1 of consensus"
46572. .47422
repeat_region /note="L1 repeat: matches 2198. .1356 of consensus"
47151. .47767
repeat_region /note="MER25 repeat: matches 2135. .1510 of consensus"
48136. .48214
repeat_region /note="MSTA repeat: matches 79. .1 of consensus"
48215. .49832
repeat_region /note="MST-INTERNAL repeat: matches 1651. .1 of consensus"
49833. .50232
repeat_region /note="MER25 repeat: matches 1504. .1297 of consensus"
50697. .50903
repeat_region /note="MER25 repeat: matches 1296. .1079 of consensus"
51475. .51694
repeat_region /note="AluXc repeat: matches 297. .1 of consensus"
51723. .52016
repeat_region /note="MER25 repeat: matches 1057. .602 of consensus"
52028. .52484
repeat_region /note="MER25 repeat: matches 548. .49 of consensus"
52672. .53166
repeat_region /note="MER25 repeat: matches 128. .11 of consensus"
53615. .53733
repeat_region /note="FLAM-A repeat: matches 128. .11 of consensus"
53875. .54925
repeat_region /note="THE1B-INTERNAL repeat: matches 107. .1167 of
consensus"
54929. .55248
repeat_region /note="THE1C repeat: matches 54. .368 of consensus"
55252. .56389
repeat_region /note="L1 repeat: matches 3226. .2078 of consensus"
56380. .56619
repeat_region /note="L1MB5 repeat: matches 242. .2 of consensus"
56472. .57179
repeat_region /note="L1 repeat: matches 5390. .4668 of consensus"
57183. .57486
repeat_region /note="AluSq repeat: matches 1. .303 of consensus"
57509. .59386

repeat_region /note="L1 repeat: matches 4680. .2773 of consensus"
59391. .59423
repeat_region /note="L1 copies of 3 mer 85 & conserved"
59435. .59746
repeat_region /note="AluJo repeat: matches 1. .302 of consensus"
59770. .59802
repeat_region /note="AluXq/g repeat: matches 36. .4 of consensus;
incomplete repeat"

Query Match 50.8%; Score 26.4; DB 9; Length 75793;
Best Local Similarity 75.0%; Pred. No. 71;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 4 aaagattccatttggaataagcaagatctgttactttgttcg 47
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27259 AAATAATACATTGGTAGGTGCTAATCTGTTACTGTATTCG 27302

RESULT 9
AC105955
LOCUS AC105955 65370 bp DNA linear HTG 11-JAN-2002
DEFINITION Mus musculus clone RP24-262D23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC105955
VERSION AC105955.1 GI:18129455
KEYWORDS HTG: HTGS, PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 65370)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-262D23
Unpublished
2 (bases 1 to 65370)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Glade,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Katat,A., Karatas,A., Kells,C., Labrecque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19374
Center clone name: 262_D_23

* NOTE: This record contains 82 individual
* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 687: contig of 687 bp in length
* 688 787: gap of 100 bp
* 788 1478: contig of 691 bp in length
* 1479 1578: gap of 100 bp
* 1579 2270: contig of 692 bp in length
* 2271 2370: gap of 100 bp
* 2371 3090: contig of 720 bp in length
* 3091 3190: gap of 100 bp
* 3191 3884: contig of 694 bp in length
* 3885 3984: gap of 100 bp
* 3985 4681: contig of 697 bp in length
* 4682 4781: gap of 100 bp
* 4782 5481: contig of 700 bp in length
* 5482 5581: gap of 100 bp
* 5582 6290: contig of 709 bp in length
* 6291 6390: gap of 100 bp
* 6391 7051: contig of 661 bp in length
* 7052 7151: gap of 100 bp
* 7152 7845: contig of 694 bp in length
* 7846 7945: gap of 100 bp
* 7946 8642: contig of 697 bp in length
* 8643 8742: gap of 100 bp
* 8743 9427: contig of 685 bp in length
* 9428 9527: gap of 100 bp
* 9528 10213: contig of 686 bp in length
* 10214 10313: gap of 100 bp
* 10314 11004: contig of 691 bp in length
* 11005 11104: gap of 100 bp
* 11105 11813: contig of 709 bp in length
* 11814 11913: gap of 100 bp
* 11914 12615: contig of 702 bp in length
* 12616 12715: gap of 100 bp
* 12716 13443: contig of 728 bp in length
* 13444 13543: gap of 100 bp
* 13544 14236: contig of 693 bp in length
* 14237 14336: gap of 100 bp
* 14337 15034: contig of 698 bp in length
* 15035 15134: gap of 100 bp
* 15135 15823: contig of 689 bp in length
* 15824 15923: gap of 100 bp
* 15924 16603: contig of 680 bp in length
* 16604 16703: gap of 100 bp
* 16704 17403: contig of 700 bp in length
* 17404 17503: gap of 100 bp
* 17504 18190: contig of 687 bp in length
* 18191 18290: gap of 100 bp
* 18291 18991: contig of 701 bp in length
* 18992 19091: gap of 100 bp
* 19092 19786: contig of 695 bp in length
* 19787 19886: gap of 100 bp
* 19887 20596: contig of 710 bp in length
* 20597 20696: gap of 100 bp
* 20697 21411: contig of 715 bp in length
* 21412 21511: gap of 100 bp
* 21512 22228: contig of 717 bp in length
* 22229 22328: gap of 100 bp
* 22329 23040: contig of 712 bp in length
* 23041 23140: gap of 100 bp
* 23141 23835: contig of 695 bp in length
* 23836 23935: gap of 100 bp
* 23936 24617: contig of 682 bp in length
* 24618 24717: gap of 100 bp
* 24718 25407: contig of 690 bp in length
* 25408 25507: gap of 100 bp

* 25508 26160: contig of 653 bp in length
* 26161 26260: gap of 100 bp
* 26261 26984: contig of 724 bp in length
* 26985 27084: gap of 100 bp
* 27085 27802: contig of 718 bp in length
* 27803 27902: gap of 100 bp
* 27903 28619: contig of 717 bp in length
* 28620 28719: gap of 100 bp
* 28720 29439: contig of 720 bp in length
* 29440 29539: gap of 100 bp
* 29540 30226: contig of 687 bp in length
* 30227 30326: gap of 100 bp
* 30327 31023: contig of 697 bp in length
* 31024 31123: gap of 100 bp
* 31124 31824: contig of 701 bp in length
* 31825 31924: gap of 100 bp
* 31925 32600: contig of 676 bp in length
* 32601 32700: gap of 100 bp
* 32701 33385: contig of 685 bp in length
* 33386 33485: gap of 100 bp
* 33486 34173: contig of 688 bp in length
* 34174 34273: gap of 100 bp
* 34274 34967: contig of 694 bp in length
* 34968 35067: gap of 100 bp
* 35068 35762: contig of 695 bp in length
* 35763 35862: gap of 100 bp
* 35863 36565: contig of 703 bp in length
* 36566 36665: gap of 100 bp
* 36666 37383: contig of 718 bp in length
* 37384 37483: gap of 100 bp
* 37484 38207: contig of 724 bp in length
* 38208 38307: gap of 100 bp
* 38308 39025: contig of 718 bp in length
* 39026 39125: gap of 100 bp
* 39126 39824: contig of 699 bp in length
* 39825 39924: gap of 100 bp
* 39925 40634: contig of 710 bp in length
* 40635 40734: gap of 100 bp
* 40735 41430: contig of 696 bp in length
* 41431 41530: gap of 100 bp
* 41531 42207: contig of 677 bp in length
* 42208 42307: gap of 100 bp
* 42308 42990: contig of 683 bp in length
* 42991 43090: gap of 100 bp
* 43091 43733: contig of 643 bp in length
* 43734 43833: gap of 100 bp
* 43834 44550: contig of 717 bp in length
* 44551 44650: gap of 100 bp
* 44651 45375: contig of 725 bp in length
* 45376 45475: gap of 100 bp
* 45476 46180: contig of 705 bp in length
* 46181 46280: gap of 100 bp
* 46281 46981: contig of 701 bp in length
* 46982 47081: gap of 100 bp
* 47082 47778: contig of 697 bp in length
* 47779 47878: gap of 100 bp
* 47879 48561: contig of 683 bp in length
* 48562 48661: gap of 100 bp
* 48662 49359: contig of 698 bp in length
* 49360 49459: gap of 100 bp
* 49460 50140: contig of 681 bp in length
* 50141 50240: gap of 100 bp
* 50241 50927: contig of 687 bp in length
* 50928 51027: gap of 100 bp
* 51028 51703: contig of 676 bp in length
* 51704 51803: gap of 100 bp
* 51804 52490: contig of 687 bp in length
* 52491 52590: gap of 100 bp
* 52591 53290: contig of 700 bp in length
* 53291 53390: gap of 100 bp
* 53391 54113: contig of 723 bp in length
* 54114 54213: gap of 100 bp
* 54214 54910: contig of 697 bp in length

```

* 54911 55010: gap of 100 bp

Query Match 50.08; Score 26; DB 2; Length 65370;
Best Local Similarity 70.08; Pred. No. 96;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 aataaagattccattggataagaacgactctgttacttgcgtcagt 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64983 AATAAATATTCATTTTITAAAAAGAAATGTTGCTTCAGTTTCTGT 65032

RESULT 10
ALI139381
LOCUS 116021 bp DNA linear PRI 12-OCT-2001
DEFINITION Human DNA sequence from clone RP11-74A12 on chromosome
13q31.2-32.2, complete sequence.
ACCESSION ALI139381
VERSION ALI139381.24 GI:16116424
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 116021)
Direct Submission
Submitted (11-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 12, 2001 this sequence version replaced gi:15990615.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em.: EMBL; Sw.:
SWISSPROT; Tr.: TrEMBL; Wp.: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-74A12 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-74A12 it may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-46412 is at 114022 in this
sequence. The true right end of clone RP11-140119 is at 2000 in
this sequence.
Location/Qualifiers
1..116021
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="q31.2-32.2"
/clone="RP11-74A12"
/clone_lib="RPCI-11.1"
complement(88879..88909)
/note="Single clone region. Assembly confirmed by
restriction digest data."

misc_feature
1 aataaagattccattggataagaacgactctgttacttgcgtt 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103282 AGAAAATATTCATTTTITAAAAAGAAATGTTGCTTCAGTTTCTGT 103326

Query Match 49.68; Score 25.8; DB 9; Length 116021;
Best Local Similarity 73.38; Pred. No. 1.1e+02;
Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 aataaagattccattggataagaacgactctgttacttgcgtt 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103282 AGAAAATATTCATTTTITAAAAAGAAATGTTGCTTCAGTTTCTGT 103326

RESULT 11
AC008771
LOCUS 123169 bp DNA linear PRI 28-FEB-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2015H6, complete sequence.
ACCESSION AC008771
VERSION AC008771.4 GI:13162500
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 123169)
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 123169)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
2 (bases 1 to 123169)
Unpublished
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 123169)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (28-FEB-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Feb 28, 2001 this sequence version replaced gi:7709299.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.3% of Sequence;
Estimated Total Number of Errors is 0.9.
STS Content:
SHGC-B4992 G53847
WI-15051 G23648
WI-17790 G24283.
Location/Qualifiers
1..123169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2015H6"
BASE COUNT 35484 a 23783 c 24111 g 39791 t
ORIGIN

Query Match 49.68; Score 25.8; DB 9; Length 123169;
Best Local Similarity 73.38; Pred. No. 1.1e+02;
Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 aataaagattccattggataagaacgactctgttacttgcgtt 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79746 AGTAAAGATCCCTTGGATGATATATCTGTTATTATTGATT 79790

RESULT 12
AC018764/c

```

```

LOCUS       AC018764               126054 bp    DNA    linear    PRI 09-NOV-2000
DEFINITION  Homo sapiens chromosome 5 clone CTD-232715, complete sequence.
ACCESSION   AC018764
VERSION     AC018764.6  GI:11128366
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 126054)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 126054)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (19-DEC-1999) Production Sequencing Facility, DOE Joint
             Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 126054)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (01-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
             Drive, Walnut Creek, CA 94598, USA
REFERENCE   4 (bases 1 to 126054)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (09-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
             Drive, Walnut Creek, CA 94598, USA
COMMENT     On Nov 9, 2000 this sequence version replaced gi:8886981.
             Draft Sequence Produced by DOE Joint Genome Institute
             www.jgi.doe.gov
             Finishing Completed at Stanford Human Genome Center
             www.shgc.stanford.edu
             Quality: Phrap Quality >=40 99.5% of Sequence;
             Estimated Total Number of Errors is 0.5.
             STS Content:
             WI-17790 G24283
             WI-15051 G23648
             Bases 89328 to 90656 excised as IS:Tn10.

FEATURES             source
     source           1..126054
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-232715"
BASE COUNT  38749 a 24943 c 24689 g 37673 t
ORIGIN
Query Match 49.6%; Score 25.8; DB 9; Length 126054;
Best Local Similarity 73.3%; Pred. No. 1.1e+02;
Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 aataaaagattccatttgatagaagaacgaactcttactttgctt 45
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 101816 AGTAAAGATTCCCTTGGTCATATACGTGATTATTTTGATT 101772

RESULT 13
AC090825/c
LOCUS       AC090825               143867 bp    DNA    linear    HTG 18-NOV-2001
DEFINITION  Homo sapiens chromosome 15 clone CTD-2054N24 map 15, WORKING DRAFT
             SEQUENCE, 3 unordered pieces.
ACCESSION   AC090825
VERSION     AC090825.3  GI:16974176
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 143867)
AUTHORS     Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE       Homo sapiens chromosome 15, clone CTD-2054N24

```

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 143867)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A.,
Camrata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodgson, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galaqan, J.,
Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
Lehoczky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

COMMENT

Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 18, 2001 this sequence version replaced gi:14336507.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12840
Center clone name: 2054_N_24
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 143329 bases at least Q40
Consensus quality: 143488 bases at least Q30
Consensus quality: 143566 bases at least Q20
Insert size: 172000; agarose-fp
Insert size: 143667; sum-of-contigs
Quality coverage: 10.5 in Q20 bases; agarose-fp
Quality coverage: 12.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 24723: contig of 24723 bp in length
24724 24823: gap of 100 bp
24824 46565: contig of 21742 bp in length
46566 46665: gap of 100 bp
46666 143867: contig of 97202 bp in length.

FEATURES
source

1..143867
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="CTD-2054N24"
/clone_lib="CITD Human BAC"
1..24723
/note="assembly_fragment"

misc_feature

misc_feature 24824..46565
/note="assembly_fragment"
misc_feature 46666..143867
/note="assembly_fragment"
BASE COUNT 45409 a 27723 c 27453 g 43082 t 200 others
ORIGIN

Query Match 49.6%; Score 25.8; DB 2; Length 143867;
Best Local Similarity 73.3%; Pred. No. 1e+02;
Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 aataaagattccattggataagaacgaatctgttacttgcctt 45
II III IIII IIII IIII IIII IIII IIII IIII
Db 94755 AAAAAATTTTCTGCTGGATGAGAACGATCTGCTACTGTTT 94711

RESULT 14
AC096094
LOCUS
DEFINITION Rattus norvegicus clone CH230-24H1, *** SEQUENCING IN PROGRESS ***,
62 unordered pieces.
AC096094
AC096094.2 GI:17943777
HTG: HTGS-PHASE1.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 176476)

REFERENCE
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,K., Blimie,K., Blomquist,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dedrich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Roife,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmari,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 176476)

AUTHORS
TITLE
JOURNAL

COMMENT

Worley,K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15627714.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GEKO
Center clone name: CH230-24H1
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 140728 bases at least Q40
Consensus quality: 150414 bases at least Q30
Consensus quality: 157122 bases at least Q20
Estimated insert size: 149244; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1
4911: contig of 4911 bp in length
5011: gap of unknown length
10561: contig of 5550 bp in length
10661: gap of unknown length
13934: contig of 3273 bp in length
14034: gap of unknown length
20595: contig of 6561 bp in length
20596: gap of unknown length
27180: contig of 6485 bp in length
27181: gap of unknown length
33208: contig of 5928 bp in length
33209: gap of unknown length
37842: contig of 4534 bp in length
37942: gap of unknown length
41481: contig of 3539 bp in length
41581: gap of unknown length
47156: contig of 5575 bp in length
47256: gap of unknown length
51520: contig of 4264 bp in length
51521: gap of unknown length
54846: contig of 3226 bp in length
54946: gap of unknown length
58569: contig of 3623 bp in length
58669: gap of unknown length
61873: contig of 3204 bp in length
61973: gap of unknown length
66179: contig of 4206 bp in length
66279: gap of unknown length
71150: contig of 4871 bp in length
71251: gap of unknown length
75339: contig of 4089 bp in length
75439: gap of unknown length
79077: contig of 3638 bp in length
79078: gap of unknown length
79178: contig of 3231 bp in length
82408: gap of unknown length
82508: gap of unknown length
87001: contig of 4493 bp in length
87101: gap of unknown length
91317: contig of 4216 bp in length

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 48313: contig of 48313 bp in length
 * 48314 48413: gap of 100 bp
 * 48414 50035: contig of 1622 bp in length
 * 50036 50135: gap of 100 bp
 * 50136 54638: contig of 4503 bp in length
 * 54639 54738: gap of 100 bp
 * 54739 58868: contig of 4130 bp in length
 * 58869 58968: gap of 100 bp
 * 58969 64007: contig of 5039 bp in length
 * 64008 64107: gap of 100 bp
 * 64108 71220: contig of 7113 bp in length
 * 71221 71320: gap of 100 bp
 * 71321 76136: contig of 4816 bp in length
 * 76137 76236: gap of 100 bp
 * 76237 82769: contig of 6533 bp in length
 * 82770 82869: gap of 100 bp
 * 82870 104915: contig of 22046 bp in length
 * 104916 105015: gap of 100 bp
 * 105016 119090: contig of 14075 bp in length
 * 119091 119190: gap of 100 bp
 * 119191 167682: contig of 48492 bp in length
 * 167683 167782: gap of 100 bp
 * 167783 179762: contig of 11980 bp in length.

FEATURES

source
 1. .179762
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RP11-115014"
 /clone="RP11-115014" Human Male BAC"
 misc_feature
 1. .48313
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 48414..50035
 /note="assembly_fragment"
 misc_feature
 50136..54638
 /note="assembly_fragment"
 misc_feature
 54739..58868
 /note="assembly_fragment"
 misc_feature
 58969..64007
 /note="assembly_fragment"
 misc_feature
 64108..71220
 /note="assembly_fragment"
 misc_feature
 71321..76136
 /note="assembly_fragment"
 misc_feature
 76237..82769
 /note="assembly_fragment"
 misc_feature
 82870..104915
 /note="assembly_fragment"
 misc_feature
 105016..119090
 /note="assembly_fragment"
 misc_feature
 119191..167682
 /note="assembly_fragment"
 misc_feature
 167783..179762
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right"

BASE COUNT 52681 a 37316 c 36703 g 51962 t 1100 others
 ORIGIN

Query Match 49.6%; Score 25.8; DB 2; Length 179762;
 Best Local Similarity 73.3%; Pred. No. le+02;
 Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 aataaaagattccattgataagaacgaactgttacttgcctt 45
 |||||
 Db 159094 AAAAAATTTTCTCTGGTGGATGAGAACGGATCTGGTACTGTGTT 159138

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: June 19, 2002, 17:14:11 ; Search time 3798.65 Seconds
(without alignments)
23.503 Million cell updates/sec

Title: US-09-462-955B-1_COPY_734_785

Perfect score: 52

Sequence: 1 aataaagattccattgga.....tggtactttgcttgagtcg 52

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1980.DAT.*
2: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1981.DAT.*
3: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1982.DAT.*
4: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1983.DAT.*
5: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1984.DAT.*
6: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1985.DAT.*
7: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1986.DAT.*
8: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1987.DAT.*
9: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1988.DAT.*
10: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1989.DAT.*
11: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1990.DAT.*
12: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1991.DAT.*
13: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1992.DAT.*
14: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1993.DAT.*
15: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1994.DAT.*
16: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1995.DAT.*
17: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1996.DAT.*
18: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1997.DAT.*
19: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1998.DAT.*
20: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA1999.DAT.*
21: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2000.DAT.*
22: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001A.DAT.*
23: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2001B.DAT.*
24: /net/abss06/SIDS1/gcgdata/hold-geneq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	26.4	50.8	243	21	Promoter from inte
C 2	25.4	48.8	6198	22	Human ovarian and
C 3	25.4	48.8	6198	22	Human reproductive
C 4	25	48.1	811	22	Human CDNA clone (
C 5	24.6	47.3	18765	24	Human glutamate re
C 6	24.6	47.3	1082138	21	Arabidopsis thalia
C 7	24.4	46.9	2763	22	Human CDNA sequenc
C 8	24.2	46.5	1173	22	Plasmid pBNpnt ins
C 9	24.2	46.5	15366	22	Genomic sequence #

C 10	23.4	45.0	575	22	AAH70492	Human cervical can
C 11	23.4	45.0	1113	22	AAH73180	Human cervical can
C 12	23.4	45.0	20835	22	AAK86765	Human immune/haema
C 13	23.4	45.0	25806	22	AAK86766	Human immune/haema
C 14	23.4	45.0	580073	18	AAH58840	Mycoplasma genital
C 15	23.2	44.6	2756	22	AAH14861	Human CDNA sequenc
C 16	23.2	44.6	6305	16	AAQ96127	Tobacco leaf antif
C 17	23	44.2	403	24	AAH17049	Human protein kina
C 18	23	44.2	578	21	AAH16113	Human colon cancer
C 19	23	44.2	732	21	AAH97370	Human colorectal c
C 20	23	44.2	1079	22	AAH60801	Human cancer agent
C 21	23	44.2	1931	21	AAH16149	Human prostate can
C 22	22.8	43.8	338	21	AAH01250	Human secreted pro
C 23	22.8	43.8	459	22	AAK83834	Human immune/haema
C 24	22.8	43.8	459	22	AAK83835	Human immune/haema
C 25	22.8	43.8	546	23	AAH78220	DNA encoding novel
C 26	22.8	43.8	790	22	AAH08051	Human CDNA clone (
C 27	22.8	43.8	1164	22	AAH161292	Human polynucleoti
C 28	22.8	43.8	1238	22	AAH192942	Human polynucleoti
C 29	22.8	43.8	1473	22	AAH159506	Human polynucleoti
C 30	22.8	43.8	2459	23	ABL26096	Drosophila melanog
C 31	22.8	43.8	2506	22	AAH15133	Human CDNA sequenc
C 32	22.8	43.8	3206	22	AAH58337	Human GTP-binding
C 33	22.8	43.8	8043	16	AAQ85924	Human protein tyro
C 34	22.8	43.8	8119	21	AAH61840	DNA encoding a hum
C 35	22.8	43.8	16449	23	ABL25556	Drosophila melanog
C 36	22.8	43.8	465237	24	ABA90193	Human oestrogen re
C 37	22.6	43.5	429	22	AAH53000	S. epidermidis ope
C 38	22.6	43.5	604	20	AAH10665	cDNA encoding a hu
C 39	22.6	43.5	2718	22	AAH160849	Human polynucleoti
C 40	22.6	43.5	3038	20	AAH234199	Human PRO541 nucle
C 41	22.6	43.5	3038	21	AAH78550	Human PRO541 (UNQ3
C 42	22.6	43.5	3038	22	AAH45959	Human DNA encoding
C 43	22.6	43.5	3145	22	AAH54834	S. epidermidis gen
C 44	22.6	43.5	3771	22	AAH54702	S. epidermidis gen
C 45	22.6	43.5	6850	24	AAH61365	Human gene regulat

ALIGNMENTS

RESULT 1
AAA96835/c
ID AAA96835 standard; DNA; 243 BP.

XX AAA96835;

XX 19-FEB-2001 (first entry)

XX Promoter from intergenic region of Commelina yellow mottle virus.

XX Promoter; intergenic region; Commelina yellow mottle virus;

XX chimeric expression promoter; plant vascular expression promoter;

XX plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; ss.

XX Commelina yellow mottle virus.

XX WO200058485-A1.

XX 05-OCT-2000.

XX 29-MAR-2000; 2000WO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

XX (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

XX WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-488786/53.

New isolated ovarian and/or breast cancer related nucleic acids and polypeptides, useful for diagnosing, treating and/or preventing human diseases and disorders, particularly ovarian and/or breast cancer -

Disclosure; SEQ ID NO 535; 577pp + Sequence Listing; English.

The invention relates to novel genes (ABA07454-ABA080224) and proteins (ABB10743-ABB10980) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 6198 BP; 1979 A; 1100 C; 1215 G; 1904 T; 0 other;

Query Match 48.8%; Score 25.4; DB 22; Length 6198;
 Best Local Similarity 74.4%; Pred. NO. 11;
 Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 6 aagattccatttggaagacgaacatctgttactttgttgcga 48
 ||| ||||| ||| ||||| ||||| ||||| |||||

Db 4704 atgacagcgtttggattaaaacatttctgttactttgcatgca 4746

RESULT 3

AAAL06924

ID AAAL06924 standard; DNA; 6198 BP.

XX AC AAAL06924;

XX XX

DT 21-NOV-2001 (first entry)

XX Human reproductive system related antigen DNA SEQ ID NO: 9612.

XX Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy; ds.
 KW Homo sapiens.

OS WO200155320-A2.

PN 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01339.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

01-SEP-2000; 2000US-0229287.
 01-SEP-2000; 2000US-0229343.
 01-SEP-2000; 2000US-0229344.
 01-SEP-2000; 2000US-0229345.
 05-SEP-2000; 2000US-0229509.
 05-SEP-2000; 2000US-0229513.
 06-SEP-2000; 2000US-0230437.
 06-SEP-2000; 2000US-0230438.
 08-SEP-2000; 2000US-0231242.
 08-SEP-2000; 2000US-0231243.
 08-SEP-2000; 2000US-0231244.
 08-SEP-2000; 2000US-0231413.
 08-SEP-2000; 2000US-0231414.
 08-SEP-2000; 2000US-0232080.
 08-SEP-2000; 2000US-0232081.
 12-SEP-2000; 2000US-0231968.
 14-SEP-2000; 2000US-0232397.
 14-SEP-2000; 2000US-0232398.
 14-SEP-2000; 2000US-0232399.
 14-SEP-2000; 2000US-0232400.
 14-SEP-2000; 2000US-0232401.
 14-SEP-2000; 2000US-0233063.
 14-SEP-2000; 2000US-0233064.
 14-SEP-2000; 2000US-0233065.
 21-SEP-2000; 2000US-0234223.
 21-SEP-2000; 2000US-0234274.
 25-SEP-2000; 2000US-0234997.
 25-SEP-2000; 2000US-0234998.
 26-SEP-2000; 2000US-0235484.
 27-SEP-2000; 2000US-0235834.
 27-SEP-2000; 2000US-0235836.
 29-SEP-2000; 2000US-0236327.
 29-SEP-2000; 2000US-0236367.
 29-SEP-2000; 2000US-0236368.
 29-SEP-2000; 2000US-0236369.
 29-SEP-2000; 2000US-0236370.
 02-OCT-2000; 2000US-0236802.
 02-OCT-2000; 2000US-0237037.
 02-OCT-2000; 2000US-0237038.
 02-OCT-2000; 2000US-0237039.
 13-OCT-2000; 2000US-0239935.
 13-OCT-2000; 2000US-0239937.
 20-OCT-2000; 2000US-0240960.
 20-OCT-2000; 2000US-0241221.
 20-OCT-2000; 2000US-0241785.
 20-OCT-2000; 2000US-0241786.
 20-OCT-2000; 2000US-0241787.
 20-OCT-2000; 2000US-0241808.
 20-OCT-2000; 2000US-0241809.
 20-OCT-2000; 2000US-0241826.
 01-NOV-2000; 2000US-0244617.
 08-NOV-2000; 2000US-0246474.
 08-NOV-2000; 2000US-0246475.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246478.
 08-NOV-2000; 2000US-0246523.
 08-NOV-2000; 2000US-0246524.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246609.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 17-NOV-2000; 2000US-0249207.
 17-NOV-2000; 2000US-0249208.
 17-NOV-2000; 2000US-0249209.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249212.
 17-NOV-2000; 2000US-0249213.
 17-NOV-2000; 2000US-0249214.
 17-NOV-2000; 2000US-0249215.
 17-NOV-2000; 2000US-0249216.
 17-NOV-2000; 2000US-0249217.
 17-NOV-2000; 2000US-0249218.
 17-NOV-2000; 2000US-0249244.
 17-NOV-2000; 2000US-0249245.
 17-NOV-2000; 2000US-0249264.
 17-NOV-2000; 2000US-0249265.
 17-NOV-2000; 2000US-0249297.
 17-NOV-2000; 2000US-0249299.
 17-NOV-2000; 2000US-0249300.
 01-DEC-2000; 2000US-0250160.
 01-DEC-2000; 2000US-0250391.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 05-JAN-2001; 2001US-02559678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-465570/50.
 Isolated nucleic acid molecule encoding a reproductive system antigen -
 is used in preventing, treating or ameliorating a medical condition -
 Disclosure; SEQ ID NO 9612; 1297pp + Sequence Listing; English.
 The present invention provides the protein and coding sequences of a
 number of human reproductive system related antigens. These can be used
 in the prevention and treatment of reproductive system disorders,
 including cancer. The present sequence is a genomic sequence encoding a
 protein of the invention.
 Sequence 6198 BP; 1979 A; 1100 C; 1215 G; 1904 T; 0 other;
 Query Match 48.8%; Score 25.4; DB 22; Length 6198;
 Best Local Similarity 74.4%; Pred. No. 11;
 Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 6 aagattccattgttgataagaacgaatctgttactttgttgcga 48
 ||| ||||| ||| ||||| ||||| |||||
 Db 4704 atgacagcgtttgttgataaacaacattctgttactttgtgcga 4746
 RESULT 4
 AAH04923
 ID AAH04923 standard; cDNA; 811 BP.
 XX
 AC AAH04923;
 XX
 DT 26-JUN-2001 (first entry)
 XX Human cDNA clone (5'-primer) SEQ ID NO:1758.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 XX
 PN EF1074617-A2.
 XX


```

XX (ARRI-) ARRIS PHARM CORP.
XX PA
XX Estell DA, Fisher J, Hartman C, McGrath M, Schmidt B;
XX PI
XX WPI; 1997-259023/23.
XX DR P-PSDB; AAW26365.
XX DR
XX Modified subtilisin inhibitor - having altered reactive site
PT resulting in decreased immunogenicity, used in detergent
PT composition(s)
XX
XX Example 1; Fig 12; 69pp; English.
XX
XX This DNA sequence comprises a DNA insert in vector pBN3 comprising
CC the Bacillus subtilis aprE promoter and the bacillus
CC amyloquelaciens subtilisin gene signal sequence, pro-sequence and
CC terminator region (see AAW26365). The vector, pBNpt, can be used
CC for the expression in B. subtilis of novel subtilisin inhibitors in
CC which a human standard mechanism inhibitor, such as pancreatic
CC secretory trypsin inhibitor, has at least one amino acid of the
CC reactive site substituted so that its dissociation constant for
CC inhibition of subtilisin is reduced by at least a factor of 100.
CC The novel subtilisin inhibitors (see AAW26362-64) combine the low
CC allergenicity of human standard mechanism inhibitors, which are not
CC specific for subtilisin, and the high affinity of non-human
CC subtilisin inhibitors such as Streptomyces subtilisin inhibitor and
CC turkey mucoid third domain protein. They form a complex with
CC subtilisin, so as to avoid problems of allergenicity in detergent
CC formulations, and can also be used in affinity purification and
CC (diagnostic) quantification of subtilisin.
XX
XX Sequence 1173 BP; 358 A; 249 C; 230 G; 336 T; 0 other;
SQ

Query Match 46.5%; Score 24.2; DB 18; Length 1173;
Best Local Similarity 78.4%; Pred. No. 23;
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 13 catttgataagaacgaatctgttacttctgttcag 49
Db 55 CGTTGGAAATCAGAGTCTGTATTATTGTATGACG 19

RESULT 9
AAS28623/c
ID AAS28623 standard; DNA; 15366 BP.
AC AAS28623;
XX
XX 07-NOV-2001 (first entry)
XX
XX Genomic sequence #463 encoding for novel human respiratory antigen.
XX Human; respiratory antigen; respiratory disorder; throat disorder;
KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
KW anti allergic; anti asthmatic; anti inflammatory; olfactory;
KW respiratory active; ds.
XX
XX Homo sapiens.
XX
XX WO20015448-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-0501333.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184564.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX
18-APR-2000; 2000US-0198123.
19-MAY-2000; 2000US-0205515.
07-JUN-2000; 2000US-0209467.
28-JUN-2000; 2000US-0214886.
30-JUN-2000; 2000US-0215135.
07-JUL-2000; 2000US-0216647.
07-JUL-2000; 2000US-0216880.
11-JUL-2000; 2000US-0217487.
11-JUL-2000; 2000US-0217496.
14-JUL-2000; 2000US-0218290.
26-JUL-2000; 2000US-0220963.
26-JUL-2000; 2000US-0220964.
14-AUG-2000; 2000US-0224518.
14-AUG-2000; 2000US-0224519.
14-AUG-2000; 2000US-0225213.
14-AUG-2000; 2000US-0225214.
14-AUG-2000; 2000US-0225266.
14-AUG-2000; 2000US-0225267.
14-AUG-2000; 2000US-0225268.
14-AUG-2000; 2000US-0225270.
14-AUG-2000; 2000US-0225447.
14-AUG-2000; 2000US-0225757.
14-AUG-2000; 2000US-0225758.
14-AUG-2000; 2000US-0225759.
18-AUG-2000; 2000US-0226279.
22-AUG-2000; 2000US-0226681.
22-AUG-2000; 2000US-0226868.
22-AUG-2000; 2000US-0227182.
23-AUG-2000; 2000US-0227009.
30-AUG-2000; 2000US-0228924.
01-SEP-2000; 2000US-0229287.
01-SEP-2000; 2000US-0229343.
01-SEP-2000; 2000US-0229344.
01-SEP-2000; 2000US-0229345.
05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233065.
21-SEP-2000; 2000US-0234223.
21-SEP-2000; 2000US-0234274.
25-SEP-2000; 2000US-0234997.
25-SEP-2000; 2000US-0234998.
26-SEP-2000; 2000US-0235484.
27-SEP-2000; 2000US-0235834.
27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
02-OCT-2000; 2000US-0237037.
02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
02-OCT-2000; 2000US-0237040.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.

```


Query Match 45.0%; Score 23.4; DB 22; Length 575;
Best Local Similarity 81.8%; Pred. No. 39;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 10 ttccatttgataagaacgaatctgttactttg 42
| | | | | | | | | | | | | | | | | | | | | |
Db 392 TCCTTTTGGATAAAACGATATCTGTGCTTGTG 360

RESULT 11
AAH73180/c
ID AAH73180 standard; cDNA; 1113 BP.

XX AC

XX AAH73180;

DT 19-SEP-2001 (first entry)

XX Human cervical cancer marker nucleic acid 4454.

XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX Homo sapiens.

XX WO200142467-A2.

XX 14-JUN-2001.

XX 08-DEC-2000; 2000WO-US33312.

XX 08-DEC-1999; 99US-0169681.

XX 21-DEC-1999; 99US-0171350.

XX 14-MAR-2000; 2000US-0189315.

XX 12-MAY-2000; 2000US-0203791.

XX 09-JUN-2000; 2000US-0210600.

XX 21-JUL-2000; 2000US-0220114.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Deeds J, Berger A, Zhao X;

XX WPI; 2001-375006/39.

XX New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -

XX Claim 1; Page 974; 1051pp; English.

XX The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.

XX Sequence 1113 BP; 358 A; 172 C; 223 G; 347 T; 13 other;

Query Match 45.0%; Score 23.4; DB 22; Length 1113;
Best Local Similarity 81.8%; Pred. No. 44;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 10 ttccatttgataagaacgaatctgttactttg 42
| | | | | | | | | | | | | | | | | | | | | |
Db 774 TCCTTTTGGATAAAACGATATCTGTGCTTGTG 742

RESULT 12

AAK86765/c

ID AAK86765 standard; DNA; 20835 BP.

XX AC AAK86765;
XX 07-NOV-2001 (first entry)
DT Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41577.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX Cytostatic; gene therapy; vaccine; metastasis; ds.
KW Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226688.
XX 22-AUG-2000; 2000US-0227182.
XX 30-AUG-2000; 2000US-0227009.
XX 01-SEP-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 06-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0232081.
XX 14-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.

```
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-024617.
PR 01-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Disclosure; SEQ ID NO 41577; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 20835 BP; 4759 A; 5358 C; 5249 G; 5469 T; 0 other;
XX
Query Match 45.0%; Score 23.4; DB 22; Length 20835;
Best Local Similarity 67.3%; Pred. No. 75;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 aataaagattccatttgataagaacgaatctgttacttgcgtgcaag 49
||| | ||||| | ||||| || ||| |||||
DB 19017 AACATAATTCATTAGTTAAGAAATTTTCAGTTCTGCTTGTCAG 18969
||| | ||||| | ||||| || ||| |||||
RESULT 13
AAK86766/c
ID AAK86766 standard; DNA; 25806 BP.
XX
XX AAK86766;
AC
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:41578.
KW Human; immune; haematopoietic; immune/hematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01354.
XX PF
XX 31-JAN-2000; 2000US-0179065.
PR
```

PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220563.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236800.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR (HUMA-) HUMAN GENOME SCI INC.
PR Rosen CA, Barash SC, Ruben SM;
PR WPI; 2001-483426/52.
PR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PR useful for preventing, diagnosing and/or treating cancers and
PR metastasis -
PR XX

FT	percentage identity to glycerol-3-phosphate dehydrogenase (GUT2) from <i>S. cerevisiae</i> "	
FT	49377..49643	
FT	/*tag= o	
FT	/label= MG041	
FT	/note= "The encoded protein shows 48.86 percentage identity to phosphoinositidinoprotein-hexose phosphotransferase (ptsH) from <i>Mycoplasma capricolum</i> "	
FT	50060..51520	
FT	/*tag= p	
FT	/label= MG042	
FT	/note= "Previously identified as MORF-19832 and MORF-20108, the encoded protein shows 41.92 percentage identity to spermidine/putrescine transport Atp-binding protein (potA) from <i>E. coli</i> "	
FT	51525..52382	
FT	/*tag= q	
FT	/label= MG043	
FT	/note= "Previously identified as MORF-20110, the encoded protein shows 26.51 percentage identity to spermidine/putrescine transport system permease protein (potB) from <i>E. coli</i> "	
FT	52366..53220	
FT	/*tag= r	
FT	/label= MG044	
FT	/note= "Previously identified as MORF-20111, the encoded protein shows 29.45 percentage identity to spermidine/putrescine transport system permease protein C (potC) from <i>E. coli</i> "	
FT	54658..55605	
FT	/*tag= s	
FT	/label= MG046	
FT	/note= "Previously identified as MORF-20112, the encoded protein shows 36.60 percentage identity to sialoglycoprotease (gcp) from <i>Pasteurella haemolytica</i> "	
FT	complement (56970..58310)	
FT	/*tag= t	
FT	/label= MG048	
FT	/note= "Previously identified as MORF-19834, MORF-20114 and MORF-20115, the encoded protein shows 43.02 percentage identity to signal recognition particle protein (ffh) from <i>B. subtilis</i> "	
FT	58117..59079	
FT	/*tag= u	
FT	/label= MG049	
FT	/note= "Previously identified as MORF-20114 and MORF-20115, the encoded protein shows 44.78 percentage identity to purine-nucleoside phosphorylase (deob) from <i>E. coli</i> "	
FT	59083..59734	
FT	/*tag= v	
FT	/label= MG050	
FT	/note= "Previously identified as MORF-20117, the encoded protein shows 83.03 percentage identity to deoxyribose-phosphate aldolase (deoc) from <i>Mycoplasma pneumoniae</i> "	
FT	complement (64898..65731)	
FT	/*tag= w	
FT	/label= MG056	
FT	/note= "Previously identified as MORF-20122, the encoded protein shows 30.25 percent identity to the protein disclosed in GB:D26185_99 from <i>B. subtilis</i> "	
FT	complement (65713..66249)	
FT	/*tag= x	
FT	/label= MG057	
FT	/note= "Previously identified as MORF-20123, the encoded protein shows 38.90 percentage identity to the protein disclosed in	

```

FT      CDS                               GB:D26185_104 from B. subtilis"
FT      1047...82597
FT      /*tag= y
FT      /label= MG067
FT      /note= "Previously identified as MORF-19845, the
FT      encoded protein shows 28.84 percentage
FT      identity to glutamic acid specific protease
FT      (SPase) from Staphylococcus aureus"
FT      91065...91919
FT      /*tag= z
FT      /label= MG070
FT      /note= "Previously identified as MORF-20136, the
FT      encoded protein shows 34.8 percentage
FT      identity to ribosomal protein S2 (rps2)
FT      from Spirulina plantensis"

Query Match      45.0%; Score 23.4; DB 18; Length 580073;
Best Local Similarity 67.3%; Pred. No. 1.3e+02;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0

QY      4 aaaaatttcatttgataagaacgaattcttactttgttcgagtcg 52
      |||| | ||||| || | ||| | |||| | ||||
Db 438256 aaatacaccatttgattgtgatacaattttcttgcgttcg 438304

RESULT 15
AAH14861
ID AAH14861 standard; cDNA; 2756 BP.
XX
AC AAH14861;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:12702.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
WPI; 2001-318749/34.
XX
Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -
XX
Claim 8; SEQ ID 12702; 2537pp + CD ROM; English.
XX
The present invention describes primer sets for synthesising 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end

```

CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 2756 BP; 911 A; 438 C; 542 G; 865 T; 0 other;

Query Match 44.6%; Score 23.2; DB 22; Length 2756;
 Best Local Similarity 77.8%; Pred. No. 61;
 Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tccatttgatgaagaacgaatctgttactttgcttg 46
 ||||| ||||| ||||| ||||| |||||
 Db 1747 tccatttgatcagaacacgtatgttccattgcttg 1782

Search completed: June 19, 2002, 17:15:17
 Job time: 18975 sec

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	52	100.0	1291	4	US-09-462-975-1	Sequence 1, Appli
2	23.2	44.6	6305	4	US-08-687-580B-6	Sequence 6, Appli
3	23	44.2	403	4	US-08-387-212-2	Sequence 2, Appli
c	23	44.2	578	4	US-04-385-982-118	Sequence 118, App
	22.8	43.8	8040	1	US-08-596-291-1	Sequence 1, Appli
	22.8	43.8	8040	3	US-09-100-804-1	Sequence 1, Appli
6	22.8	43.8	8043	3	PCT-US94-09943-1	Sequence 1, Appli
7	22.8	43.8	8043	5	US-09-290-640-45	Sequence 45, Appli
8	22.8	43.8	8119	4	US-08-341-587-7	Sequence 7, Appli
9	22	42.3	28720	4	US-08-474-067-1	Sequence 1, Appli
c	20.6	41.5	3959	1	US-08-474-068A-1	Sequence 1, Appli
10	21.6	41.5	3959	2	US-08-472-481-1	Sequence 1, Appli
11	21.6	41.5	3959	2	US-09-268-992-7	Sequence 7, Appli
12	21.6	41.5	72604	4	US-09-103-663-35	Sequence 35, Appli
c	21.4	41.2	1344	3	US-08-742-185-101	Sequence 101, App
13	21.4	41.2	2738	4	US-08-936-165A-33	Sequence 33, Appli
14	21.4	41.2	2738	4	US-08-920-812-12	Sequence 12, Appli
15	21.2	40.8	557	4	US-08-920-827-12	Sequence 12, Appli
16	21.2	40.8	557	4	US-08-921-177-12	Sequence 12, Appli
c	21.2	40.8	4500	1	US-08-362-577C-12	Sequence 12, Appli
17	21.2	40.8	4500	1	US-08-920-828-12	Sequence 12, Appli
18	21.2	40.8	4500	1	US-08-553-619B-6	Sequence 6, Appli
19	21.2	40.8	4550	4		
20	21.2	40.8	4795	3		
c	21	40.4	423	4		
21	21	39.6	2441	1		
22	20.6	39.6	2441	1		
c	20.6	39.6	2441	1		
23	20.6	39.6	2441	1		
c	20.6	39.6	2441	1		
24	20.6	39.6	2441	1		
c	20.6	39.6	2441	1		
25	20.6	39.6	2441	1		
26	20.6	39.6	2441	2		
c	20.4	39.2	2173	2		
27	20.4	39.2	2173	2		

0:

NAME	REGISTRATION NUMBER	REFERENCE/DOCKET NUMBER	TELECOMMUNICATION INFORMATION	TELEPHONE	TELEFAX	TELEX	INFORMATION FOR SEQ ID NO: 1	SEQUENCE CHARACTERISTICS	LENGTH	TYPE	STRANDEDNESS	TOPOLOGY	MOLECULE TYPE	HYPOTHETICAL	ANTI-SENSE	ORIGINAL SOURCE	ORGANISM	FEATURE	NAME/KEY	LOCATION	US-08-596-291-1	Query Match	Best Local Similarity	Matches	Score	DB	Length	Indels	Mismatches	Gaps	
GATES, EDWARD R.	31,616	LO461/7000		617/720-3500	617/720-2441	92-1742 EZEKIEL	1:	8040 base pairs	nucleic acid	single	linear	cdna to mRNA	NO	NO		HOMO SAPIENS			CDS	78..7475		43.8%	71.4%	30;	22.8;	1;	8040;	12;	0;	0;	
<p>US-08-596-291-1</p> <p>Query Match 43.8%; Score 22.8; DB 1; Length 8040;</p> <p>Best Local Similarity 71.4%; Pred. No. 11;</p> <p>Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;</p>																															
<p>QY 6 aagattccattgtgataagcaacgaatctgttacttgcgtgc 47</p> <p> </p> <p>Db 6822 AAGATACACGTTGGGAAGAGAGAGCTGCTTTACATTGCCCTGC 6863</p>																															
<p>RESULT 6</p> <p>US-09-100-804-1</p> <p>; Sequence 1, Application US/09100804</p> <p>; Patent No. 6066472</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: GONEZ, LEONEL JORGE</p> <p>; APPLICANT: SARAS, JAN</p> <p>; APPLICANT: CLAESON-WELSH, LENA</p> <p>; APPLICANT: HELDIN, CARL-HENRIK</p> <p>; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL</p> <p>; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN</p> <p>; TITLE OF INVENTION: TYROSINE PHOSPHATASES</p> <p>; NUMBER OF SEQUENCES: 34</p> <p>; CORRESPONDENCE ADDRESS:</p> <p>; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.</p> <p>; STREET: 600 ATLANTIC AVENUE</p> <p>; CITY: BOSTON</p> <p>; STATE: MASSACHUSETTS</p> <p>; COUNTRY: USA</p> <p>; ZIP: 02210</p> <p>; COMPUTER READABLE FORM:</p> <p>; MEDIUM TYPE: Floppy disk</p> <p>; COMPUTER: IBM PC compatible</p> <p>; OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>; SOFTWARE: PatentIn Release #1.0, Version #1.25</p> <p>; CURRENT APPLICATION DATA:</p> <p>; APPLICATION NUMBER: US/09/100,804</p> <p>; FILING DATE:</p> <p>; CLASSIFICATION:</p> <p>; PRIORITY APPLICATION DATA:</p> <p>; APPLICATION NUMBER: US 08/596,291</p> <p>; FILING DATE: 09-AUG-1996</p> <p>; APPLICATION NUMBER: US 08/115,573</p> <p>; FILING DATE: 01-SEP-1993</p> <p>; PRIORITY APPLICATION DATA:</p> <p>; APPLICATION NUMBER: PCT/US94/09943</p> <p>; FILING DATE: 01-SEP-1994</p> <p>; ATTORNEY/AGENT INFORMATION:</p> <p>; NAME: GATES, EDWARD R.</p> <p>; REGISTRATION NUMBER: 31,616</p>																															

```
;
; REFERENCE/DOCKET NUMBER: LO461/7003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8040 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 78..7475
; PCT-US94-09943-1
;
; Query Match 43.8%; Score 22.8; DB 3; Length 8040;
; Best Local Similarity 71.4%; Pred. No. 11;
; Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
;
; QY 6 aagattccattgtgataagaacgaatctgttacttctgttc 47
;      ||||| ||||| ||||| ||||| ||||| ||||| |||||
; DB 6822 AAGATACCAAGTTGGGAAAGAGAGTTCGTTTACATTCCTGTC 6863
;
; RESULT 7
; PCT-US94-09943-1
; Sequence 1, Application PC/TUS9409943
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09943
; FILING DATE: 01-SEP-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/115,573
; FILING DATE: 01-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: TWOMEY, MICHAEL J.
; REGISTRATION NUMBER: P-38,349
; REFERENCE/DOCKET NUMBER: LO461/7000WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; TELEX: 92-1742 EZEKIEL
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8043 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 78..7478
; PCT-US94-09943-1
;
; Query Match 43.8%; Score 22.8; DB 4; Length 8119;
; Best Local Similarity 71.4%; Pred. No. 11;
; Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
;
; QY 6 aagattccattgtgataagaacgaatctgttacttctgttc 47
;      ||||| ||||| ||||| ||||| ||||| ||||| |||||
; DB 6868 aagataccagttgggaagaagagtttcgtttacattgctgc 6909
;
; RESULT 8
; US-09-290-640-45
; Sequence 45, Application US/09290640
; Patent No. 6204055
; GENERAL INFORMATION:
; APPLICANT: Dean, Nicholas M.
; APPLICANT: Marcussion, Eric G.
; TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling
; FILE REFERENCE: ISPH-0351
; CURRENT APPLICATION NUMBER: US/09/290,640
; CURRENT FILING DATE: 1999-04-12
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 45
; LENGTH: 8119
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64)..(7521)
; PUBLICATION INFORMATION:
; JOURNAL: FEBS Lett.
; VOLUME: 337
; ISSUE: 2
; PAGES: 200-206
; DATE: 1994-01-10
; DATABASE ACCESSION NUMBER: D21209/Genbank
; DATABASE ENTRY DATE: 1999-02-05
; US-09-290-640-45
;
; Query Match 43.8%; Score 22.8; DB 4; Length 8119;
; Best Local Similarity 71.4%; Pred. No. 11;
; Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
;
; QY 6 aagattccattgtgataagaacgaatctgttacttctgttc 47
;      ||||| ||||| ||||| ||||| ||||| ||||| |||||
; DB 6868 aagataccagttgggaagaagagtttcgtttacattgctgc 6909
;
; RESULT 9
; US-09-341-587-7/c
; Sequence 7, Application US/09341587
; Patent No. 6346606
; GENERAL INFORMATION:
; APPLICANT: Mollenhauer, Jan
; TITLE OF INVENTION: Protein Containing an SRCR Domain
; FILE REFERENCE: 4121-108
; CURRENT APPLICATION NUMBER: US/09/341,587
; CURRENT FILING DATE: 1999-08-31
; EARLIER APPLICATION NUMBER: PCT/DE98/00096
```

GENERAL INFORMATION:
APPLICANT: Ranscht, Barbara
TITLE OF INVENTION: T-Cadherin Adhesion Molecule
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700

```

; FEATURE:
; NAME/KEY: modified base
; LOCATION: all n positions
; OTHER INFORMATION: n-a, c, g, or t
US-09-268-992-7

Query Match 41.5%; Score 21.6; DB
Best Local Similarity 68.2%; Pred. No. 45;
Matches 30; Conservative 0; Mismatches

QY 1 aataaagattccattgggataagaacgaatctgttactt
   || ||| ||| | ||| ||| ||| ||| |||
Db 12430 aaataaataataaataaataaataaataatgaatgtttcctt

RESULT 14
US-09-190-965-2/c
; Sequence 2, Application US/09190965
; Patent No. 6071721
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Gorgone, Gina A.
; TITLE OF INVENTION: CALCIUM BINDING PROTEIN
; FILE REFERENCE: PF-0635 US
; CURRENT APPLICATION NUMBER: US/09/190,965
; CURRENT FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 3734805
US-09-190-965-2

Query Match 41.2%; Score 21.4; DB
Best Local Similarity 71.8%; Pred. No. 26;
Matches 28; Conservative 0; Mismatches

QY 13 catttggataagaacgaatctgttactttgttgcaagt
   | |||| | | ||| | | ||| ||| ||| |||
Db 295 CGTTTGtACCACACAGAATTTCTTTCAATTGCTTGCAGT

RESULT 15
US-08-795-868-17
; Sequence 17, Application US/08795868
; Patent No. 5846773
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: A SINGLE GENE ENCODING A
; NUMBER OF INVENTIONS: AND STRIATED-SPECIFIC MU
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,868

```



```

; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/494,577
; FILING DATE: 22-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/032001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-795-868-17

```

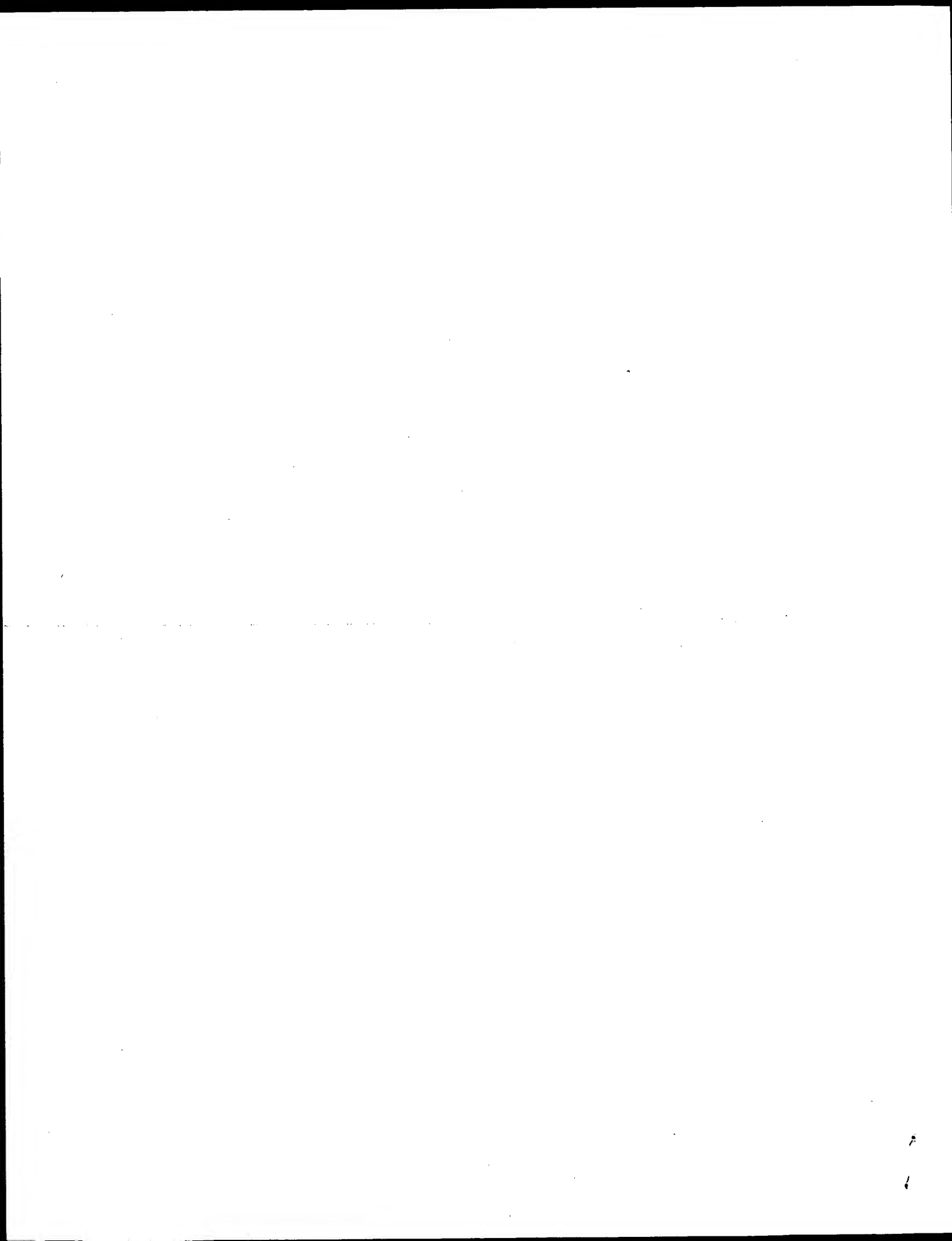
```

Query Match          41.2%; Score 21.4; DB 2; Length 2738;
Best Local Similarity 66.0%; Pred. No. 30;
Matches 31; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2 ataaaagattccattggataagaacgaactctgttacttgccttgca 48
   ||| |||| | |||| | |||| | |||| | |||| | |||| | |||
Db 489 ATATAAGAGCATTGTGAAAAAAACTAATGTTGAAATCGCTGGCA 535

```

Search completed: June 19, 2002, 16:33:08
Job time: 16446 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 14:04:31 ; Search time 7489.97 seconds
(without alignments)
93.704 Million cell updates/sec

Title: US-09-462-955B-1_COPY_734_785

Perfect score: 52

Sequence: 1 aataaaagattccattgga.....tggtactttgttcgagtcgc 52

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB ID	Description
1	26.4	50.8	257	12	AQ106136
2	26.2	50.4	999	10	BG749262
3	26	50.0	864	12	A2547272
4	25.8	49.6	418	9	A1065614
5	25.8	49.6	487	12	B43802
6	25.8	49.6	537	12	AQ694110
7	25.4	48.8	525	12	A2390369
8	25.4	48.8	576	10	BG654372
9	25.4	48.8	781	10	BE884333
10	25	48.1	433	10	BE678889
11	25	48.1	456	10	BE678999
12	25	48.1	487	12	AQ270104
13	25	48.1	501	10	BE678820
14	25	48.1	560	10	BE678247
15	25	48.1	625	12	AQ020281
16	25	48.1	811	9	AU122315
17	25	48.1	904	12	CNS049BD

C 18	24.8	47.7	577	12	AQ506286
C 19	24.8	47.7	648	12	AG160388
C 20	24.6	47.3	516	10	BM198468
C 21	24.6	47.3	713	12	AG089691
C 22	24.4	46.9	321	9	AA221297
C 23	24.4	46.9	365	10	BF547464
C 24	24.4	46.9	367	10	BF016895
C 25	24.4	46.9	385	10	BE688287
C 26	24.4	46.9	390	9	AI021393
C 27	24.4	46.9	412	12	AZ010087
C 28	24.4	46.9	421	9	BB675436
C 29	24.4	46.9	650	9	AV344344
C 30	24.4	46.9	708	10	BI157582
C 31	24.4	46.9	747	12	CNS0324A
C 32	24.2	46.5	317	10	BF747486
C 33	24.2	46.5	375	9	AA228092
C 34	24.2	46.5	394	9	AW037099
C 35	24.2	46.5	398	12	AQ135393
C 36	24.2	46.5	418	10	BF728680
C 37	24.2	46.5	507	9	AW566513
C 38	24.2	46.5	517	9	AV634321
C 39	24.2	46.5	565	10	BM382402
C 40	24.2	46.5	626	10	BM380636
C 41	24.2	46.5	628	10	BG842500
C 42	24.2	46.5	650	9	AW303369
C 43	24.2	46.5	711	10	BF630600
C 44	24.2	46.5	1101	12	CNS00KH0
C 45	24	46.2	312	10	BF062027

ALIGNMENTS

RESULT 1
AQ106136
LOCUS
DEFINITION HS_3055_A2_B07_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3055 Col=14 Row=C, DNA sequence.
ACCESSION AQ106136
VERSION AQ106136.1 GI:3481492
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 257)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3055 row: C column: 14
Class: BAC ends
High quality sequence stop: 257.
Location/Qualifiers
1. 257
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3055 Col=14 Row=C"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"

```

BASE COUNT      93 a   30 c   43 g   91 t
ORIGIN

Query Match      50.8%; Score 26.4; DB 12; Length 257;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 4 aaagattccatttgataagaacgaatctgttacttgcgttc 47
||||| ||||| ||| | ||||| ||||| ||||| |||||
Db 48 AAAATAATACATTGGGTAGTGTCTAATCTGTACTGTATTGTC 91

RESULT 2
BG749262/c
LOCUS
DEFINITION 602708092F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4844615 5',
mRNA sequence.
ACCESSION BG749262
VERSION BG749262.1 GI:14059915
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 999)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1682 row: h column: 24
High quality sequence stop: 810.
FEATURES
Location/Qualifiers
1..999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4844615"
/clone_lib="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGCAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. |"
BASE COUNT      213 a   267 c   311 g   208 t
ORIGIN

Query Match      50.4%; Score 26.2; DB 10; Length 999;
Best Local Similarity 79.5%; Pred. No. 58;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 8 gattccatttgataagaacgaatctgttacttgcgttc 46
||||| ||||| ||| | ||||| ||||| ||||| |||||
Db 589 GATTCCATCTGTTATGAGGAGTGTACCACCTTTCCTTC 551

RESULT 3
AZ547272/c
LOCUS
DEFINITION ENT0220TR Entamoeba histolytica Sheared DNA Entamoeba histolytica

genomic, DNA sequence.
ACCESSION AZ547272
VERSION AZ547272.1 GI:11169813
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 864)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 16
High quality sequence stop: 811.
Location/Qualifiers
1..864
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pPOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      352 a   103 c   90 g   319 t
ORIGIN

Query Match      50.0%; Score 26; DB 12; Length 864;
Best Local Similarity 70.0%; Pred. No. 67;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 aataaagattccatttgataagaacgaatctgttacttgcgttc 50
||||| ||||| ||| | ||||| ||||| ||||| |||||
Db 604 AATAAAGATATTATTGGTTAAATTTGAATTTGGTAATATTGTTGTTAT 555

RESULT 4
AI065614/c
LOCUS
DEFINITION ag90e12.x1 maize inflorescence immature ear library Zea mays cDNA
clone ag90e12 3', mRNA sequence.
ACCESSION AI065614
VERSION AI065614.1 GI:3341021
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 416)
AUTHORS Schut, K., de la Bastide, M., Gnoj, L., Habermann, K., Huang, E. N.,
Parnell, L. D., Dedhia, N., Martienssen, R. and McCombie, W. R.

```

<div><div>TITLE</div><div>Expressed sequence tags from Z. mays</div><div>Unpublished (1998)</div><div>CONTACT: W. Richard McCombie</div><div>Liita Annenberg Hazen Genome Sequencing Center</div><div>Cold Spring Harbor Laboratory</div><div>PO Box 100, Cold Spring Harbor, NY 11724, USA</div><div>Tel: 516 367 8884</div><div>Fax: 516 367 8874</div><div>Email: mcombie@cshl.org</div><div>Plate: ag90 row: e column: 12</div><div>Seq primer: M13 forward universal -21</div><div>High quality sequence stop: 418.</div><div>Location/Qualifiers</div><div>1. .418</div><div>/organism="Zea mays"</div><div>/cultivar="B73"</div><div>/db_xref="taxon:4577"</div><div>/clone="ag90el2"</div><div>/clone_lib="maize inflorescence immature ear library"</div><div>/sex="female"</div><div>/tissue_type="immature ear"</div><div>/note="Vector: pBLUESCRIPT SK+ (X52325); Site_1: XhoI; Site_2: EcoRI; This library is described in Schmidt, Hake, et al., (1993) Plant Cell 5:729-737. cDNAs are directionally cloned into the XhoI and EcoRI sites; XhoI is near the polyA tail. Most reads from this library are 3' in direction. Additional information on this library as well as ftp access to all sequences can be found at http://www.cshl.org/maizegenome"</div></div>	<div><div>BASE COUNT</div><div>121 a 98 c 78 g 121 t</div><div>ORIGIN</div><div>Query Match 49.6%; Score 25.8; DB 9; Length 418;</div><div>Best Local Similarity 73.3%; Pred. No. 73;</div><div>Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;</div></div> <div><div>QY</div><div>2</div><div>ataaagaattccatttgataagaacgaatctgttactttgcttg 46</div><div> </div><div>Db 362 ACAAAAGCTTGGATTGGAGATAGACCTATCTGCTGCTTGCCTG 318</div></div> <div><div>RESULT</div><div>5</div><div>B43802</div><div>LOCUS</div><div>HS-1058-A2-D06-MR.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate-CT 780 Col-12 Row-G, DNA sequence.</div><div>ACCESSION</div><div>B43802</div><div>VERSION</div><div>B43802.1</div><div>GI:2548636</div><div>KEYWORDS</div><div>GSS.</div><div>SOURCE</div><div>human.</div><div>ORGANISM</div><div>Homo sapiens</div><div>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</div><div>REFERENCE</div><div>1 (bases 1 to 487)</div><div>AUTHORS</div><div>Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.</div><div>TITLE</div><div>Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors</div><div>JOURNAL</div><div>Unpublished (1997)</div><div>COMMENT</div><div>Contact: Mahairas GG, Zackrone KD, Hood L University of Washington Seattle, WA 98195, USA Tel: (206) 616-8744 Fax: (206) 685-7301 Email: kzackrone@u.washington.edu Sequence Tagged Connector Plate: CT 780 row: G column: 12 Class: BAC ends High quality sequence stop: 487.</div></div> <div><div>FEATURES</div><div>source</div><div>1. .487</div><div>Location/Qualifiers</div></div>	<div><div>TITLE</div><div>Expressed sequence tags from Z. mays</div><div>Unpublished (1998)</div><div>CONTACT: W. Richard McCombie</div><div>Liita Annenberg Hazen Genome Sequencing Center</div><div>Cold Spring Harbor Laboratory</div><div>PO Box 100, Cold Spring Harbor, NY 11724, USA</div><div>Tel: 516 367 8884</div><div>Fax: 516 367 8874</div><div>Email: mcombie@cshl.org</div><div>Plate: ag90 row: e column: 12</div><div>Seq primer: M13 forward universal -21</div><div>High quality sequence stop: 418.</div><div>Location/Qualifiers</div><div>1. .418</div><div>/organism="Zea mays"</div><div>/cultivar="B73"</div><div>/db_xref="taxon:4577"</div><div>/clone="ag90el2"</div><div>/clone_lib="maize inflorescence immature ear library"</div><div>/sex="female"</div><div>/tissue_type="immature ear"</div><div>/note="Vector: pBLUESCRIPT SK+ (X52325); Site_1: XhoI; Site_2: EcoRI; This library is described in Schmidt, Hake, et al., (1993) Plant Cell 5:729-737. cDNAs are directionally cloned into the XhoI and EcoRI sites; XhoI is near the polyA tail. Most reads from this library are 3' in direction. Additional information on this library as well as ftp access to all sequences can be found at http://www.cshl.org/maizegenome"</div></div>	<div><div>BASE COUNT</div><div>121 a 98 c 78 g 121 t</div><div>ORIGIN</div><div>Query Match 49.6%; Score 25.8; DB 9; Length 418;</div><div>Best Local Similarity 73.3%; Pred. No. 73;</div><div>Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;</div></div> <div><div>QY</div><div>2</div><div>ataaagaattccatttgataagaacgaatctgttactttgcttg 46</div><div> </div><div>Db 362 ACAAAAGCTTGGATTGGAGATAGACCTATCTGCTGCTTGCCTG 318</div></div> <div><div>RESULT</div><div>5</div><div>B43802</div><div>LOCUS</div><div>HS-1058-A2-D06-MR.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate-CT 780 Col-12 Row-G, DNA sequence.</div><div>ACCESSION</div><div>B43802</div><div>VERSION</div><div>B43802.1</div><div>GI:2548636</div><div>KEYWORDS</div><div>GSS.</div><div>SOURCE</div><div>human.</div><div>ORGANISM</div><div>Homo sapiens</div><div>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</div><div>REFERENCE</div><div>1 (bases 1 to 487)</div><div>AUTHORS</div><div>Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.</div><div>TITLE</div><div>Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors</div><div>JOURNAL</div><div>Unpublished (1997)</div><div>COMMENT</div><div>Contact: Mahairas GG, Zackrone KD, Hood L University of Washington Seattle, WA 98195, USA Tel: (206) 616-8744 Fax: (206) 685-7301 Email: kzackrone@u.washington.edu Sequence Tagged Connector Plate: CT 780 row: G column: 12 Class: BAC ends High quality sequence stop: 487.</div></div> <div><div>FEATURES</div><div>source</div><div>1. .487</div><div>Location/Qualifiers</div></div>	<div><div>Expressed sequence tags from Z. mays</div><div>Unpublished (1998)</div><div>CONTACT: W. Richard McCombie</div><div>Liita Annenberg Hazen Genome Sequencing Center</div><div>Cold Spring Harbor Laboratory</div><div>PO Box 100, Cold Spring Harbor, NY 11724, USA</div><div>Tel: 516 367 8884</div><div>Fax: 516 367 8874</div><div>Email: mcombie@cshl.org</div><div>Plate: ag90 row: e column: 12</div><div>Seq primer: M13 forward universal -21</div><div>High quality sequence stop: 418.</div><div>Location/Qualifiers</div><div>1. .418</div><div>/organism="Zea mays"</div><div>/cultivar="B73"</div><div>/db_xref="taxon:4577"</div><div>/clone="ag90el2"</div><div>/clone_lib="maize inflorescence immature ear library"</div><div>/sex="female"</div><div>/tissue_type="immature ear"</div><div>/note="Vector: pBLUESCRIPT SK+ (X52325); Site_1: XhoI; Site_2: EcoRI; This library is described in Schmidt, Hake, et al., (1993) Plant Cell 5:729-737. cDNAs are directionally cloned into the XhoI and EcoRI sites; XhoI is near the polyA tail. Most reads from this library are 3' in direction. Additional information on this library as well as ftp access to all sequences can be found at http://www.cshl.org/maizegenome"</div></div>	<div><div>BASE COUNT</div><div>121 a 98 c 78 g 121 t</div><div>ORIGIN</div><div>Query Match 49.6%; Score 25.8; DB 9; Length 418;</div><div>Best Local Similarity 73.3%; Pred. No. 73;</div><div>Matches 33; Conservative 0; Mismatches 12; Indels 0; Gaps 0;</div></div> <div><div>QY</div><div>2</div><div>ataaagaattccatttgataagaacgaatctgttactttgcttg 46</div><div> </div><div>Db 362 ACAAAAGCTTGGATTGGAGATAGACCTATCTGCTGCTTGCCTG 318</div></div> <div><div>RESULT</div><div>5</div><div>B43802</div><div>LOCUS</div><div>HS-1058-A2-D06-MR.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate-CT 780 Col-12 Row-G, DNA sequence.</div><div>ACCESSION</div><div>B43802</div><div>VERSION</div><div>B43802.1</div><div>GI:2548636</div><div>KEYWORDS</div><div>GSS.</div><div>SOURCE</div><div>human.</div><div>ORGANISM</div><div>Homo sapiens</div><div>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</div><div>REFERENCE</div><div>1 (bases 1 to 487)</div><div>AUTHORS</div><div>Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.</div><div>TITLE</div><div>Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors</div><div>JOURNAL</div><div>Unpublished (1997)</div><div>COMMENT</div><div>Contact: Mahairas GG, Zackrone KD, Hood L University of Washington Seattle, WA 98195, USA Tel: (206) 616-8744 Fax: (206) 685-7301 Email: kzackrone@u.washington.edu Sequence Tagged Connector Plate: CT 780 row: G column: 12 Class: BAC ends High quality sequence stop: 487.</div></div> <div><div>FEATURES</div><div>source</div><div>1. .487</div><div>Location/Qualifiers</div></div>	<div><div>Expressed sequence tags from Z. mays</div><div>Unpublished (1998)</div><div>CONTACT: W. Richard McCombie</div><div>Liita Annenberg Hazen Genome Sequencing Center</div><div>Cold Spring Harbor Laboratory</div><div>PO Box 100, Cold Spring Harbor, NY 11724, USA</div><div>Tel: 516 367 8884</div><div>Fax: 516 367 8874</div><div>Email: mcombie@cshl.org</div><div>Plate: ag90 row: e column: 12</div><div>Seq primer: M13 forward universal -21</div><div>High quality sequence stop: 418.</div><div>Location/Qualifiers</div><div>1. .418</div><div>/organism="Zea mays"</div><div>/cultivar="B73"</div><div>/db_xref="taxon:4577"</div><div>/clone="ag90el2"</div><div>/clone_lib="maize inflorescence immature ear library"</div><div>/sex="female"</div><div>/tissue_type="immature ear"</div><div>/note="Vector: pBLUESCRIPT SK+ (X52325); Site_1: XhoI; Site_2: EcoRI; This library is described in Schmidt, Hake, et al., (1993) Plant Cell 5:729-737. cDNAs are directionally cloned into the XhoI and EcoRI sites; XhoI is near the polyA tail. Most reads from this library are 3' in direction. Additional information on this library as well as ftp access to all sequences can be found at http://www.cshl.org/maizegenome"</div></div>	<div><div>BASE COUNT</div><div>121 a 98 c 78 g 121 t</div><div>ORIGIN</div><div>Query Match 49.6%; Score 25.8; DB 9; Length 418;</div><div>Best Local Similarity 73.3%; Pred. No. 73;</div><div>Matches 33</div></div>
---	--	---	--	---	--	---	--

DB 484 ATAAACACCCGAATTGCATATAAAGACCCCTATGACCTTTGTTGTCACAGTAC 434

RESULT 8
 BG654372 576 bp mRNA linear EST 05-JUL-2001
 LOCUS ib39d06.y1 HR85 islet Homo sapiens cDNA 5', mRNA sequence.
 DEFINITION BG654372
 ACCESSION BG654372.1 GI:13791781
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 576)
 AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
 Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
 Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blisstein,A.,
 Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
 ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
 Jackson,Y. and Bowers,Y.
 TITLE Endocrine Pancreas Consortium
 JOURNAL Unpublished (2000)
 COMMENT Other_ESTs: ib39d06.x1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Dr. Hiroshi Inoue
 (hinoue@im.wustl.edu)
 Seq primer: -40RP from Gibco
 High quality sequence stop: 429.
 Location/Qualifiers
 1..576
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HR85 islet"
 /tissue_type="Purified pancreatic islet"
 /lab_host="PH10B"
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
 NotI; Site_2: XhoI; cDNA made by oligo-dT priming. Size_
 Size_selected on agarose gel. Average insert size ~1kb. 5'
 XhoI site was destroyed after directional cloning.
 Amplified once. Contact information: Hiroshi Inoue, MD,
 Metabolism Div. (Alan Permutt Lab), Washington University
 School of Medicine, Box 8127, 660 South Euclid Ave., St.
 Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, tel:
 314-362-1916, Fax: 314-747-2692."
 BASE COUNT 172 a 106 c 96 g 202 t
 ORIGIN

Query Match 48.8%; Score 25.4; DB 10; Length 576;
 Best Local Similarity 74.4%; Pred. No. le+02;
 Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 6 aagattccattggataaagaacgaatctgttacttggcttga 48
 ||| ||||| ||| ||||| ||||| ||||| ||||| |||||
 DB 153 ATGACAGCGTTTGGATTAACACATTTCTGTACTTTGCATGCA 195

RESULT 9
 BE884333 781 bp mRNA linear EST 20-OCT-2000
 LOCUS 601505608F1 NTH_MGC_71 Homo sapiens cDNA clone IMAGE:3907441 5',
 DEFINITION mRNA sequence.
 ACCESSION BE884333

Thu Jun 20 06:56:57 2002

```

VERSION      BE884333.1  GI:103333109
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS      National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE        Unpublished (1999)
JOURNAL
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Invitro Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLM9717 row: p column: 02
              High quality sequence stop: 629.
              Location/Qualifiers
FEATURES
    source
        1..781
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3907441"
            /clone_lib="NIH_MGC_71"
            /tissue_type="leiomyosarcoma"
            /lab_host="pHL08 (phage-resistant)"
            /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
              Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
              Average insert size 2.1 kb."
              Average insert size 2.1 kb.
BASE COUNT   234 a 145 c 150 g 251 t 1 others
ORIGIN

Query Match      48.8%; Score 25.4; DB 10; Length 781;
Best Local Similarity 74.4%; Pred. No. 1.1e+02;
Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY  6 aaagatccattgataagaacgaatctgttacttctgttcgca 48
    | | | | | | | | | | | | | | | | | | | | | | | |
Db  102 ATGACACCGTTGGATTAAACATCTTCGTACTTGTGATGCA 144

RESULT 10
BE678889
LOCUS      df57a09.x1 Xenopus laevis unfertilized egg cDNA library Xenopus
DEFINITION laevis cDNA clone IMAGE:3743416 3', mRNA sequence.
ACCESSION  BE678889
VERSION     BE678889.1  GI:10061016
KEYWORDS    EST.
SOURCE      African clawed frog.
ORGANISM    Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE   1 (bases 1 to 433)
AUTHORS     Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            , B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
            WashU Xenopus EST project, 1999
            Unpublished (1999)
            Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Bruce Blumberg
            Library normalized by Jihwan Song

Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: http://www.resgen.com/ Please reference the id listed
below when ordering this clone: Source lab clone id - xlnneg002m17
Seq primer: -40UP from Gibco
High quality sequence stop: 423.
Location/Qualifiers
    1..433
        /organism="Xenopus laevis"
        /db_xref="taxon:8355"
        /clone="IMAGE:3743416"
        /clone_lib="Xenopus laevis unfertilized egg cDNA library"
        /tissue_type="unfertilized egg"
        /lab_host="top-10 F."
        /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
          XhoI; cDNA was prepared from 2ug of poly A+ RNA.
          EcoRI-XhoI cut cDNA was then ligated into UniZap-XR
          (Stratagene) with EcoRI at the 5' end and XhoI at the 3'
          end. SS-library phagemids were prepared by mass excision
          from the original library and normalized by hybridization
          to biotinylated driver (prepared from the same library by
          PCR) to Cot-omega of 11. After removal of hybrids and
          excess driver by streptavidin sepharose chromatography,
          the ss-phagemids were made double stranded and
          electroporated into Top-10 F'. Original library
          constructed by Bruce Blumberg (Blumberg et al., 1991
          Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,
          2923-2935). Note: This is a Xenopus Gene Collection (XGC)
          library."
BASE COUNT      81 a 116 c 83 g 153 t
ORIGIN

Query Match      48.1%; Score 25; DB 10; Length 433;
Best Local Similarity 69.4%; Pred. No. 1.4e+02;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY  4 aaagatccattgataagaacgaatctgttacttctgttcgagtc 52
    | | | | | | | | | | | | | | | | | | | | | | | |
Db  206 AAAGGTTGTTATTCGCAAGCAAGAAACACAGATATTTGCTTGCAATGC 254

RESULT 11
BE678999
LOCUS      df78a09.x1 Xenopus laevis oocyte non normalized Xenopus laevis cDNA
DEFINITION clone IMAGE:3745432 3', mRNA sequence.
ACCESSION  BE678999
VERSION     BE678999.1  GI:10061236
KEYWORDS    EST.
SOURCE      African clawed frog.
ORGANISM    Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE   1 (bases 1 to 456)
AUTHORS     Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            , B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
            WashU Xenopus EST project, 1999
            Unpublished (1999)
            Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Bruce Blumberg
            Library normalized by Jihwan Song

```

FEATURES

/note=Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2ug of poly A+ RNA. EcoRI-XhoI cut cDNA was then ligated into Uniap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. ss-library phagemids were prepared by mass excision from the original library and normalized by hybridization to biotinylated driver (prepared from the same library by PCR) to Cot-omega of 11. After removal of hybrids and excess driver by streptavidin sepharose chromatography, the ss-phagemids were made double stranded and electroporated into Top-10 F+. Original library constructed by Bruce Blumberg (Blumberg et al., 1991 Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9, 2923-2935). Note: This is a Xenopus Gene Collection (XGC) library."

BASE COUNT	ORIGIN	86 a	122 c	81 g	167 t

Query Match	48.1%	Score 25;	DB 10;	Length 456;
Best Local Similarity	69.4%	Pred. No.	1.4e+02;	

4 aaagattccatttggaataagaacgaatctgttacttgcagtcg 52
|||||
206 AAAGGTTGTTATTGGCAAGGAAAAACACAGATATTTTCTTGCACATCG 254

RESULT 12	AQ270104	487 bp	DNA	linear	GSS 03-NOV-1998
Q270104/c	HS-2060_AI_E04_T7	Cit	Approved Human Genomic	Sperm Library D Homo	
OCUS	sapiens	genomic clone	Plate=2060 Col=7	Row=I,	DNA sequence.
DEFINITION	AQ270104				
CCFSSSTON	AQ270104				

A2270104
 AQ270104.1 GI:3824174
 GSS.
 human.
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 487)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel.: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2060 row: I column: 7
Class: BAC ends

High quality sequence stop: 487.
 Location/Qualifiers
 1. .487

source	1. .487
BASE COUNT	171 a 82 c 78 g 154 t 2 others
	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="plate-2060 Col-7 Row-1" /clone_lib="CIT Approved Human Genomic Sperm Library D" /sex="male" /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"

Query Match	48.1%;	Score 25;	DB 12;	Length 487;
Best Local Similarity	69.4%;	Pred. No. 1.4e+02;		

Qy
2 ataaaagattccatttggataagaacgaatctgtactttgcttgcaagt 50
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db
298 AAAGATGGTTACATTAGAAAAGATCGATTCCCTTCCCTTTCCTACATT 250

RESULT 13

Accession	Gene	Length	Source	EST
AF067882	LOCUS	501 bp	linear	EST 18-APR-2001
AF067882	DEFINITION			
AF067882	df56a09.x1 Xenopus laevis unfertilized egg cDNA library			
AF067882	laevis cDNA clone IMAGE:3743320 3', mRNA sequence.			
AF067882	ACCESSION			

VERSION BE678820.1 GI:10060878
KEYWORDS EST.

SOURCE	ORGANISM
African clawed frog.	Xenopus laevis

REFERENCE
AUTHORS
Xenopodinae; Xenopus.
1 (bases 1 to 501)
Clifton, S., Johnson, S. L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person,
B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.

TITLE
JOURNAL
Washu Xenopus EST project, 1999
Unpublished (1999)
WATERSON, R. AND WILSON, R.

Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg

Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: <http://www.resgen.com/> Please reference the id listed
below when ordering this clone: Source lab clone id - xlnmeg002117
Seq primer: -400P from Gibco
High quality sequence from: 494

FEATURES

```

/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3743320"
/clone.lib="Xenopus laevis unfertilized egg cDNA library"
/tissue.type="unfertilized egg"
/lab.host="Top-10 F'"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from 2ug of poly A+ RNA.
EcoRI-XhoI cut cDNA was then ligated into UniZap-XR
(Stratagene) with EcoRI at the 5' end and XhoI at the 3',
end. SS-library phagemids were prepared by mass excision
from the original library and normalized by hybridization

```


Db 262 AAAAATGGTTACATTTAAAAAGAGATCGATTCCCTTCCTTTCTTACATT 214

Search completed: June 19, 2002, 14:04:35
Job time: 7533 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 11:59:02 ; Search time 247.21 Seconds
(without alignments)
21.860 Million cell updates/sec

Title: US-09-462-955B-1_COPY_655_676

Perfect score:

Sequence: 1 ctgccaggccgaaggcctggg 22

Scoring table: IDENTITY_NUC

IDENTITY_NOC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

```
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
```

```
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
```

```
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
```

```
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
```

5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*

```
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	22	100.0	23	4	US-09-462-975-6	Sequence 6, Appli
2	22	100.0	1291	4	US-09-462-975-1	Sequence 1, Appli
c 3	17.8	80.9	1938	4	US-09-232-200-29	Sequence 29, Appli
c 4	17.8	80.9	1938	4	US-09-232-197-29	Sequence 29, Appli
c 5	17.8	80.9	1938	4	US-09-232-201-29	Sequence 29, Appli
c 6	17.8	80.9	1941	4	US-09-232-200-28	Sequence 28, Appli
c 7	17.8	80.9	1941	4	US-09-232-197-28	Sequence 28, Appli
c 8	17.8	80.9	1941	4	US-09-232-201-28	Sequence 28, Appli
c 9	17.8	80.9	2816	1	US-08-785-241-1	Sequence 1, Appli
c 10	17.8	80.9	3217	4	US-09-232-200-64	Sequence 64, Appli
c 11	17.8	80.9	3217	4	US-09-232-197-64	Sequence 64, Appli
c 12	17.8	80.9	3217	4	US-09-232-201-64	Sequence 64, Appli
c 13	17.8	80.9	3694	4	US-09-232-200-46	Sequence 46, Appli
c 14	17.8	80.9	3694	4	US-09-232-197-46	Sequence 46, Appli
c 15	17.8	80.9	3694	4	US-09-232-201-46	Sequence 46, Appli
c 16	17.8	80.9	3704	4	US-09-232-200-24	Sequence 24, Appli
c 17	17.8	80.9	3704	4	US-09-232-197-24	Sequence 24, Appli
c 18	17.8	80.9	3704	4	US-09-232-201-24	Sequence 24, Appli
19	17.2	78.2	1431	4	US-09-491-772-7	Sequence 7, Appli
20	17.2	78.2	3131	3	US-09-035-648-23	Sequence 23, Appli
21	17.2	78.2	3131	4	US-09-001-951-23	Sequence 23, Appli
22	17.2	78.2	8580	4	US-09-491-772-1	Sequence 1, Appli
23	16.4	74.5	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c 24	16.2	73.6	472	2	US-08-811-949-40	Sequence 40, Appli
c 25	16.2	73.6	747	1	US-08-257-341-6	Sequence 6, Appli
c 26	16.2	73.6	779	1	US-08-133-804-3	Sequence 3, Appli
27	16.2	73.6	779	1	US-08-461-838-3	Sequence 3, Appli

ALIGNMENTS

```

RESULT      1
US-09-462-975-6
; Sequence 6, Application US/09462975
; Patent No. 6103345
; GENERAL INFORMATION:
; APPLICANT: Rohde, Wolfgang
; APPLICANT: Becker, Dieter
; APPLICANT: Randles, John W.
; APPLICANT: Hehn, Alain
; APPLICANT: Salamini, Francesco
; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER
; FILE REFERENCE: 23232.0003U1
; CURRENT APPLICATION NUMBER: US/09/462,975
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/EP98/04345
; PRIOR FILING DATE: 1998-07-13
; PRIOR APPLICATION NUMBER: 19730502.4
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence
; OTHER INFORMATION: construct
US-09-462-975-6

```

Query Match 100.0%; Score 22; DB 4; Length 23;

Qy 1 ctgcccaggccgaaggcctggg 22
|||||
Db 1 ctgcccaggccgaaggcctggg 22

RESULT 2
US-09-462-975-1
; Sequence 1, Application US/09462975
; Patent No. 6303345
; GENERAL INFORMATION:
; APPLICANT: Rohde, Wolfgang
; APPLICANT: Becker, Dieter
; APPLICANT: Randles, John W.
; APPLICANT: Hehn, Alain
; APPLICANT: Salamini, Francesco
; TITLE OF INVENTION: USE OF VIRUS DNA AS A PROMOTER

RESULT 4
US-09-232-197-29/c
; Sequence 29, Application US/092323197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.

RESULT 6
US-09-232-200-28/c
; Sequence 28, Application US/092322200A
: Patent No 6288213

US-09-232-197-28/c
 ; GENERAL INFORMATION:
 ; APPLICANT: Stahl, Andreas
 ; APPLICANT: Hirsch, David J.
 ; APPLICANT: Lodish, Harvey F.
 ; APPLICANT: Gimeno, Ruth E.
 ; APPLICANT: Tartaglia, Louis A.
 ; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
 ; FILE REFERENCE: WHI97-21p3MB
 ; CURRENT APPLICATION NUMBER: US/09/232,200A
 ; EARLIER FILING DATE: 1998-01-14
 ; EARLIER APPLICATION NUMBER: 60/071,374
 ; EARLIER FILING DATE: 1998-01-15
 ; EARLIER APPLICATION NUMBER: 60/093,491
 ; EARLIER FILING DATE: 1998-07-20
 ; EARLIER APPLICATION NUMBER: 60/110,941
 ; EARLIER FILING DATE: 1998-12-04
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 28
 ; LENGTH: 1941
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-232-200-28

Query Match 80.9%; Score 17.8; DB 4; Length 1941;
 Best Local Similarity 90.5%; Pred. No. 30;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcccaggccgaagcctggg 22
 ||||| ||||| ||||| |||||
 Db 517 TGCCAGGCGAGAGCCAGGG 497

RESULT 7

US-09-232-197-28/c
 ; Sequence 28, Application US/09232197A
 ; Patent No. 6300096
 ; GENERAL INFORMATION:
 ; APPLICANT: Stahl, Andreas
 ; APPLICANT: Hirsch, David J.
 ; APPLICANT: Lodish, Harvey F.
 ; APPLICANT: Gimeno, Ruth E.
 ; APPLICANT: Tartaglia, Louis A.
 ; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
 ; FILE REFERENCE: WHI97-21p3MA
 ; CURRENT APPLICATION NUMBER: US/09/232,197A
 ; EARLIER FILING DATE: 1999-01-14
 ; EARLIER APPLICATION NUMBER: 60/071,374
 ; EARLIER FILING DATE: 1998-01-15
 ; EARLIER APPLICATION NUMBER: 60/093,491
 ; EARLIER FILING DATE: 1998-07-20
 ; EARLIER APPLICATION NUMBER: 60/110,941
 ; EARLIER FILING DATE: 1998-12-04
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 28
 ; LENGTH: 1941
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-232-197-28

Query Match 80.9%; Score 17.8; DB 4; Length 1941;
 Best Local Similarity 90.5%; Pred. No. 30;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcccaggccgaagcctggg 22
 ||||| ||||| ||||| |||||
 Db 517 TGCCAGGCGAGAGCCAGGG 497

RESULT 8

US-09-232-201-28/c
 ; Sequence 28, Application US/09232201A
 ; Patent No. 6348321
 ; GENERAL INFORMATION:
 ; APPLICANT: Stahl, Andreas
 ; APPLICANT: Hirsch, David J.
 ; APPLICANT: Lodish, Harvey F.
 ; APPLICANT: Gimeno, Ruth E.
 ; APPLICANT: Tartaglia, Louis A.
 ; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
 ; FILE REFERENCE: WHI97-21p3MC
 ; CURRENT APPLICATION NUMBER: US/09/232,201A
 ; EARLIER FILING DATE: 1999-01-14
 ; EARLIER APPLICATION NUMBER: 60/071,374
 ; EARLIER FILING DATE: 1998-01-15
 ; EARLIER APPLICATION NUMBER: 60/093,491
 ; EARLIER FILING DATE: 1998-07-20
 ; EARLIER APPLICATION NUMBER: 60/110,941
 ; EARLIER FILING DATE: 1998-12-04
 ; NUMBER OF SEQ ID NOS: 105
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 28
 ; LENGTH: 1941
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-232-201-28

Query Match 80.9%; Score 17.8; DB 4; Length 1941;
 Best Local Similarity 90.5%; Pred. No. 30;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcccaggccgaagcctggg 22
 ||||| ||||| ||||| |||||
 Db 517 TGCCAGGCGAGAGCCAGGG 497

RESULT 9

US-08-785-241-1/c
 ; Sequence 1, Application US/08785241
 ; Patent No. 5695963
 ; GENERAL INFORMATION:
 ; APPLICANT: McKnight, Steven L.
 ; APPLICANT: Russell, David W.
 ; APPLICANT: Tian, Hui
 ; TITLE OF INVENTION: Endothelial PAS Domain Protein
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
 ; STREET: 268 BUSH STREET, SUITE 3200
 ; CITY: SAN FRANCISCO
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 94104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/785,241
 ; FILING DATE: 17-JAN-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: OSMAN, RICHARD A
 ; REGISTRATION NUMBER: 36,627
 ; REFERENCE/DOCKET NUMBER: UTS/SD:1229
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 343-4341
 ; TELEFAX: (415) 343-4342
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2816 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-785-241-1

Query Match 80.9%; Score 17.8; DB 1; Length 2816;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctgccaggccgaaggcctgg 21
||||||| |||||
Db 2784 CTGCCAGGTAGAAGGCGCTGG 2764

RESULT 10

US-09-232-200-64/c
; Sequence 64, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 3217
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-200-64

Query Match 80.9%; Score 17.8; DB 4; Length 3217;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcccaggccgaaggcctgg 22
||||||| |||||
Db 517 TGCCAGGCGAGAGGCCAGGG 497

RESULT 11

US-09-232-197-64/c
; Sequence 64, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941

; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 3217
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-197-64

Query Match 80.9%; Score 17.8; DB 4; Length 3217;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcccaggccgaaggcctgg 22
||||||| |||||
Db 517 TGCCAGGCGAGAGGCCAGGG 497

RESULT 12

US-09-232-201-64/c
; Sequence 64, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MC
; CURRENT APPLICATION NUMBER: US/09/232,201A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 3217
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-232-201-64

Query Match 80.9%; Score 17.8; DB 4; Length 3217;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgcccaggccgaaggcctgg 22
||||||| |||||
Db 517 TGCCAGGCGAGAGGCCAGGG 497

RESULT 13

US-09-232-200-46/c
; Sequence 46, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WH197-21p3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15

; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 3694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-200-46

Query Match 80.9%; Score 17.8; DB 4; Length 3694;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccaggccgaagcctggg 22
||||||| ||||| |||
Db 691 TGCCAGGCAGAGGCCAGGG 671

RESULT 14
US-09-232-197-46/c
; Sequence 46, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 3694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-197-46

Query Match 80.9%; Score 17.8; DB 4; Length 3694;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccaggccgaagcctggg 22
||||||| ||||| |||
Db 691 TGCCAGGCAGAGGCCAGGG 671

RESULT 15
US-09-232-201-46/c
; Sequence 46, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MC
; CURRENT APPLICATION NUMBER: US/09/232,201A

; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 3694
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-201-46

Query Match 80.9%; Score 17.8; DB 4; Length 3694;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccaggccgaagcctggg 22
||||||| ||||| |||
Db 691 TGCCAGGCAGAGGCCAGGG 671

Search completed: June 19, 2002, 15:47:08
Job time: 13686 sec

Thu Jun 20 06:56:45 2002

us-09-462-955b-1_copy_655_676.rni

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 11:59:02 ; Search time 7489.97 Seconds
(without alignments)
39.644 Million cell updates/sec

Title: US-09-462-955B-1_COPY_655_676
Perfect score: 22
Sequence: 1 ctgcccagggccgaagcctggg 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_htc.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_htc.*
- 12: gb_gss.*
- 13: gb_gss_hum.*
- 14: em_gss_inv.*
- 15: em_gss_pln.*
- 16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	DB	ID	Description
C 1	19.4	88.2	935	10	BG327577	BG327577	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 2	18.8	85.5	176	10	BE860673	BE860673	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 3	18.8	85.5	238	9	BB147701	BB147701	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 4	18.8	85.5	393	9	BB284196	BB284196	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 5	18.8	85.5	555	9	BB239924	BB239924	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 6	18.8	85.5	608	12	BH043080	BH043080	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 7	18.8	85.5	654	10	BI101919	BI101919	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 8	18.8	85.5	959	10	BF119646	BF119646	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 9	18.4	83.6	291	10	BI032357	BI032357	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 10	18.4	83.6	377	9	BB813476	BB813476	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 11	18.4	83.6	405	9	AW522130	AW522130	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 12	18.4	83.6	438	9	AA158436	AA158436	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 13	18.4	83.6	455	9	AW535323	AW535323	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 14	18.4	83.6	512	9	AI406571	AI406571	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 15	18.4	83.6	517	10	BF403937	BF403937	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 16	18.4	83.6	558	10	BI295549	BI295549	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 17	17.8	80.9	176	10	BI004403	BI004403	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001

C 18	17.8	80.9	243	9	AU058457	AU058457	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 19	17.8	80.9	295	9	AW877400	AW877400	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 20	17.8	80.9	313	9	AW336731	AW336731	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 21	17.8	80.9	349	10	BE385251	BE385251	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 22	17.8	80.9	370	10	BE771160	BE771160	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 23	17.8	80.9	371	10	BE372847	BE372847	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 24	17.8	80.9	372	10	BE771169	BE771169	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 25	17.8	80.9	375	10	BI342341	BI342341	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 26	17.8	80.9	377	10	BI404521	BI404521	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 27	17.8	80.9	383	10	BI024756	BI024756	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 28	17.8	80.9	384	9	AW416678	AW416678	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 29	17.8	80.9	390	10	BF087382	BF087382	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 30	17.8	80.9	392	10	AI520113	AI520113	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 31	17.8	80.9	397	9	AW408256	AW408256	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 32	17.8	80.9	401	9	AW838277	AW838277	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 33	17.8	80.9	409	10	BE897489	BE897489	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 34	17.8	80.9	418	10	BE818183	BE818183	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 35	17.8	80.9	430	9	BE014102	BE014102	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 36	17.8	80.9	433	9	BE013607	BE013607	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 37	17.8	80.9	434	9	AW403812	AW403812	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 38	17.8	80.9	435	10	BG609304	BG609304	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 39	17.8	80.9	440	10	BF078233	BF078233	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 40	17.8	80.9	442	9	AA405771	AA405771	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 41	17.8	80.9	445	10	BI401089	BI401089	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 42	17.8	80.9	454	9	AW416681	AW416681	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 43	17.8	80.9	462	10	BF775369	BF775369	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 44	17.8	80.9	471	10	BF444355	BF444355	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001
C 45	17.8	80.9	479	10	BF443589	BF443589	60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', linear EST 27-FEB-2001

ALIGNMENTS

RESULT 1

BG327577/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BG327577 935 bp mRNA linear EST 27-FEB-2001
60246566F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564192 5', mRNA sequence.

BG327577

BG327577.1 GI:13134015

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 935)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LCM1279 row: d column: 17

High quality sequence stop: 789.

Location/Qualifiers

1..935

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4564192"

/clone_lib="NIH_MGC_14"

/tissue_type="renal cell adenocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: kidney; Vector: pOTB7; Site: 1: XhoI; Site: 2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

FEATURES
source

the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)." 222 a 236 c 297 g 180 t

BASE COUNT
ORIGIN

Query Match 88.2%; Score 19.4; DB 10; Length 935;
Best Local Similarity 95.2%; Pred. No. 5.7e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ctgccagccgagcagcctgg 21
|||||
Db 910 CTGCCAGCCGAGGAGCCTGG 890

RESULT 2
LOCUS BE860673 176 bp mRNA linear EST 29-SEP-2000
DEFINITION UI-M-AHI-agv-c-04-0-UI.r1 NIH.BMAP_MCE.N Mus musculus cDNA clone
ACCESSION BE860673
VERSION BE860673.1 GI:10377856
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 176)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Reverse.

FEATURES
Source
1. 176
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AHI-agv-c-04-0-UI"
/clone_lib="NIH_BMAP_MCE_N"
/dev_stage="27-32 days"
/lab_host="pH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MCE library is a normalized library constructed from mouse cerebellum. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories."
BASE COUNT 29 a 58 c 40 g 49 t
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 176;
Best Local Similarity 90.9%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccagccgagcagcctgg 22
|||||
Db 71 CTGCTCAGCCGAGGCCCTGG 92

RESULT 3
LOCUS BB147701 238 bp mRNA linear EST 28-JUN-2000
DEFINITION BB147701 RIKEN full-length enriched, adult female vagina Mus musculus cDNA clone 9930101123 3', mRNA sequence.

ACCESSION BB147701
VERSION BB147701.1 GI:8802638
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 238)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
Source

Location/Qualifiers
1. 238
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="9930101123"
/clone_lib="RIKEN full-length enriched, adult female vagina"
/sex="female"
/tissue_type="vagina"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAAATTAATCCGCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I.

BASE COUNT 45 a 54 c 55 g 84 t

Query Match 85.5%; Score 18.8; DB 9; Length 238;
Best Local Similarity 90.9%; Pred. No. 7.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ctgccagggccgaagggctggg 22
||||| ||||| ||||| |||||
Db 97 CTGCTAGGCGGAGACCTGGG 118

RESULT 4
LOCUS BB284196 393 bp mRNA linear EST 01-AUG-2000
DEFINITION BB284196 RIKEN full-length enriched, adult retina Mus musculus cDNA
clone A930105B05 3', mRNA sequence.
ACCESSION BB284196
VERSION BB284196.1 GI:8984645
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 393)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermolabile and thermoactivation of the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2): 520-524 (1998)
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5): 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

TITLE
JOURNAL
COMMENT

19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
source

Location/Qualifiers
1. .393
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A930105B05"
/clone_lib="RIKEN full-length enriched, adult retina"
/tissue_type="retina"
/dev_stage="adult"
/lab_host="DH10B"

/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAAATTAATCCGCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. -Retina RNA was provided by Stefano Gustinich, Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA02115, USA, whose assistance we gratefully acknowledge."

BASE COUNT 78 a 112 c 96 g 107 t

Query Match 85.5%; Score 18.8; DB 9; Length 393;
Best Local Similarity 90.9%; Pred. No. 8.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccagggccgaagggctggg 22
||||| ||||| ||||| |||||
Db 252 CTGCTAGGCGGAGACCTGGG 273

RESULT 5
LOCUS BB239924 555 bp mRNA linear EST 23-OCT-2001
DEFINITION BB239924 RIKEN full-length enriched, 3 days neonate thymus Mus musculus cDNA clone A630079H08 3', mRNA sequence.
ACCESSION BB239924 GI:16355050
VERSION BB239924.2
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 555)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 555)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)

On Jul 5, 2000 this sequence version replaced gi:8924532.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

TITLE
JOURNAL
COMMENT

Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Konedda, Y., Ishikawa, T., Ozawa, K., Tanaka, F., Matsuura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamashita, I., Aizawa
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

FEATURES

Location/Qualifiers
 1. .555

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="A630079H08"
 /clone_lib="RIKEN full-length enriched, 3 days neonate
 thymus"
 /tissue_type="thymus"
 /dev_stage="3 days neonate"
 /lab_host="Dh10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 20.0 and subtraction to Rot = 459.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GAGAGAGAGATTCGAGTTAAATTAATCCCTCCCTCC
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a
 modified pBluescript KS(+) after bulk excision from Lambda
 FLC I."

BASE COUNT

118 a 155 c 140 g 141 t 1 others

ORIGIN

Query Match 85.5%; Score 18.8; DB 9; Length 555;
 Best Local Similarity 90.9%; Pred. No. 8.9e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ctgcccaggccgaagcctggg 22
 ||||| ||||| ||||| |||||
 Db 414 CTGCTCAGCGCGAGCCCTGGG 435

RESULT 6

BH043080/c

LOCUS

DEFINITION

BH043080

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end

page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 375 row: D column: 2

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. .608

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-375D2"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT

162 a 144 c 160 g 142 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 12; Length 608;

Best Local Similarity 90.9%; Pred. No. 9.1e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccaggccgaagcctggg 22

||||| ||||| ||||| |||||

Db 410 CTGCCCGAGCCGAGCTCCCTGGG 389

RESULT 7

BH101919/c

LOCUS

DEFINITION

BH101919

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 654)

BH043080 608 bp DNA linear GSS 17-JUL-2001
 RPCI-24-375D2.TJ RPCI-24 Mus musculus genomic clone RPCI-24-375D2,
 DNA sequence.

BH043080 608 bp DNA linear GSS 17-JUL-2001
 BH043080.1 GI:14824835
 GSS.
 house mouse.

ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 608)
 Zhao, S., Niemman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)

CONTACT: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end

page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 375 row: D column: 2

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. .608

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-375D2"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT

162 a 144 c 160 g 142 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 12; Length 608;

Best Local Similarity 90.9%; Pred. No. 9.1e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccaggccgaagcctggg 22

||||| ||||| ||||| |||||

Db 410 CTGCCCGAGCCGAGCTCCCTGGG 389

RESULT 8

BH101919/c

LOCUS

DEFINITION

BH101919

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 654)

BH101919 654 bp mRNA linear EST 26-JUN-2001
 602887615F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5042882
 5', mRNA sequence.

BH101919 654 bp mRNA linear EST 26-JUN-2001
 BH101919.1 GI:14552812
 EST.
 house mouse.

ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 654)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM1117 row: n column: 03
 High quality sequence stop: 654.
 Location/Qualifiers
 1. 654
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5042882"
 /clone_lib="NCI_CGAP_Kid14"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dn. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. |"
 BASE COUNT 172 a 168 c 181 g 133 t
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 654;
 Best Local Similarity 90.9%; Pred. No. 9.2e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ctgccagggccgaagcctggg 22
 |||||||||||
 Db 495 CTGCCAGGCGGAGTGCTGGG 474
 RESULT 8
 BF119646/c
 LOCUS BF119646
 DEFINITION 601757514F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3986834 5',
 mRNA sequence.
 ACCESSION BF119646
 VERSION BF119646.1 GI:10958790
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; 1 (bases 1 to 959)
 REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgaps-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM9192 row: 1 column: 03
 High quality sequence stop: 612.
 Location/Qualifiers
 1. 959
 /organism="Mus musculus"
 /strain="C57/B6"
 /db_xref="taxon:10090"
 /clone="IMAGE:3986834"
 /clone_lib="NCI_CGAP_Mam5"

FEATURES
 source

/tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dn. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"
 BASE COUNT 251 a 242 c 289 g 176 t 1 others
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 959;
 Best Local Similarity 90.9%; Pred. No. 9.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ctgccagggccgaagcctggg 22
 |||||||||||
 Db 64 CTGCCAGGCGGAGTGCTGGG 43
 RESULT 9
 BI032357
 LOCUS BI032357
 DEFINITION CM3-NN0246-030101-606-d03 NN0246 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BI032357
 VERSION BI032357.1 GI:14438983
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 291)
 REFERENCE Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM3&t2=CM3-NN0246-030101-606-d03&t3=2001-01-03&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 7
 High quality sequence stop: 227.
 Location/Qualifiers
 1. 291
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NN0246"
 /dev_stage="Adult"

FEATURES
 source

/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 67 a 87 c 80 g 57 t
 ORIGIN

Query Match 83.6%; Score 18.4; DB 10; Length 291;
Best Local Similarity 95.0%; Pred. No. 1.1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 tggccaggccgaagcctgg 21
||||| |||||||
Db 62 TGGCCAGCCGGAAGCCTGG 81

RESULT 10
BB813476
LOCUS BB813476 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus
DEFINITION musculus cDNA clone G730019E17 3', mRNA sequence.
ACCESSION BB813476
VERSION BB813476.1 GI:16986105
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 377)
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii
Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,
Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa
A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,
Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
2001)

JOURNAL UNPUBLISHED (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES
Source
1. 377
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G730019E17"
/clone_lib="RIKEN full-length enriched, lung RCB-0558 LLC
cDNA"
/tissue_type="lung"
/cell_line="RCB-0558 LLC"

BASE COUNT 90 a 102 c 91 g 94 t
ORIGIN

Query Match 83.6%; Score 18.4; DB 9; Length 377;
Best Local Similarity 95.0%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 gccacggccgaagcctgg 22
||||| |||||||
Db 107 GCGCCGCGGCAAGCCTGGG 126

RESULT 11
AW522130/c
LOCUS AW522130 405 bp mRNA linear EST 06-MAR-2000
DEFINITION UI-R-BJ0p-aga-h-02-0-UI.s1 UI-R-BJ0p Rattus norvegicus cDNA clone
UI-R-BJ0p-aga-h-02-0-UI 3', mRNA sequence.
ACCESSION AW522130
VERSION AW522130.1 GI:7164515
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 405)
Bonaldi, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized AV canal at 16.5 dpc library cDNA library preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
1. .405
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ0p-aga-h-02-0-UI"
/clone_lib="UI-R-BJ0p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0p
library is a subtracted library derived from the UI-R-AAL,
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the Not I site
and the oligo-dT track. The library was constructed as
described by Bonaldi, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG_LIB=UI-R-BJ0p
TAG_TISSUE=AV canal at 16.5 dpc
TAG_SEQ=GAACC"

```

BASE COUNT      90 a  105 c  99 g  111 t
ORIGIN

Query Match      83.6%; Score 18.4; DB 9; Length 405;
Best Local Similarity 95.0%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  3 gccaggccgaagcctggg 22
    ||| ||||| ||||| |||
Db  316 GCCGAGCCGAGAGGCTGGG 297

RESULT 12
LOCUS      AAL158436      438 bp      mRNA      linear      EST 09-MAR-1998
DEFINITION z059hl1.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
IMAGE:591237 5' similar to SW:CLP2_MOUSE Q08093 CALPONIN H2, SMOOTH
MUSCLE. ;, mRNA sequence.
ACCESSION  AAL158436
VERSION    AAL158436.1 GI:1733281
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 438)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marr,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevenson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2561 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 392.
FEATURES             Location/Qualifiers
     source           1..438
                     /organism="Homo sapiens"
                     /db_xref="GDB:4622623"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:591237"
                     /lab_host="Stratagene pancreas (#937208)"
                     /lab_host="SOLR cells (kanamycin resistant)"
     note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
     ECKR1; Site_2: XhoI; Cloned unidirectionally. Primer:
     Oligo dt. Pancreatic adenocarcinoma cell line. Average
     insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
     sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
     CTCGACTTTTCTTTTCTTTT 3'"
BASE COUNT      98 a  150 c  115 g  67 t  8 others
ORIGIN

Query Match      83.6%; Score 18.4; DB 9; Length 438;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  1 ctgccaggccgaagcctgg 21
    ||| ||||| ||||| |||
Db  263 CNGCCCGAGCCGAGACCTGG 243

RESULT 13
AW553523
LOCUS      L0228C11-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA
DEFINITION clone L0228C11 3', mRNA sequence.
ACCESSION  AW553523
VERSION    AW553523.1 GI:7198946
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 455)
AUTHORS   Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
III, Becker,K.G. and Ko,M.S.H.
Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
20381348
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: L0228 row: C column: 11
Seq primer: -21M13 Forward
High quality sequence stop: 455
POLYA=Yes.
FEATURES             Location/Qualifiers
     source           1..455
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:L0228C11-3"
                     /db_xref="taxon:10090"
                     /clone="L0228C11"
                     /dev_stage="Newborn Ovary"
                     /sex="female"
                     /lab_host="DH108"
                     /note="Vector: pSPORT1 (Gibco/BRL Life Technology);
                     Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
                     7 Newborn Ovary. The double-stranded cDNA was synthesized
                     by Gibco's kit with an Oligo(dT) primer [NotI
                     primer-adaptor from GibcoBRL]
                     [5'-pGACTAGTCTAGATCGGAGCGCCCTTTT-3'] from
                     2.56ug of total RNA . The double-stranded cDNAs were
                     treated with T4 DNA polymerase and purified by
                     ethanol-precipitation. The cDNAs were ligated to
                     Lone-linker LL-Sal3 (include SalI sequence). The cDNAs
                     were purified by phenol/chloroform and separated from
                     free linkers by Centricon 100. Then, cDNAs were amplified
                     by long-range high fidelity PCR using Takara's Ex Taq
                     polymerase. Then, the cDNAs were purified by
                     phenol/chloroform and by Centricon 100. The cDNAs were
                     digested with SalI and NotI enzymes. Then, the cDNAs were
                     size selected by Gibco's Size Fractionation Column. The
                     cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
                     vector. The DH10B E. coli host was transformed with the
                     ligation mixture by chemical method. The library was
                     constructed by Xiaohong Wang and Yulan Piao."
BASE COUNT      109 a  152 c  158 g  36 t
ORIGIN

Query Match      83.6%; Score 18.4; DB 9; Length 455;
Best Local Similarity 95.0%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  3 gccaggccgaagcctggg 22
    ||| ||||| ||||| |||
Db  265 GCCCAGGCTGAAGGCTGGG 284

RESULT 14
AW553523

```

```

AI406571/c
LOCUS       EST234857 512 bp mRNA linear EST 09-FEB-1999
DEFINITION Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
            RKIDY92 3' end, mRNA sequence.
ACCESSION   AI406571
VERSION     AI406571.1 GI:4250075
KEYWORDS    EST.
SOURCE      Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus
REFERENCE   1 (bases 1 to 512)
AUTHORS    Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE      Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat
            Gene Index
JOURNAL    Unpublished (1998)
COMMENT    Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..512
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone="RKIDY92"
                     /clone_lib="Normalized rat kidney, Bento Soares"
                     /note="Organ: kidney; Vector: pT73pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT   116 a 135 c 137 g 124 t
ORIGIN
1
2
3 gccaggccgaagcctggg 22
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized thalamus library cDNA Library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..517
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CAL-bjd-j-19-0-UI"
                     /clone_lib="UI-R-CAL"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT73D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CAL
                     library is a subtracted library derived from the following
                     tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                     , midbrain, cerebral cortex, corpus striatum, testis, and
                     hippocampus. For a detailed description of the library
                     from which this clone was derived, please visit our web
                     site at ratest.eng.uiowa.edu. The subtraction has been
                     previously described in (Bonaldo, Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_LIB=UI-R-CAL
                     TAG_TISSUE=thalamus
                     TAG_SEQ=GATCG"
BASE COUNT   114 a 134 c 133 g 136 t
ORIGIN
1
2
3 gccaggccgaagcctggg 22
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

Search completed: June 19, 2002, 14:04:08
Job time: /506 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 11:59:02 ; Search time 5904.86 Seconds
(without alignments)
77.967 Million cell updates/sec

Title: US-09-462-955B-1_COPY_655_676
Perfect score: 22
Sequence: 1 ctgccagccgaaggcctggg 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----	-----	-----	-----	-----	-----

1	22	100.0	23	6	AR172775	Sequence
2	22	100.0	1291	14	AR172770	Sequence
3	22	100.0	1291	14	CFDCG	M29563 Coconut fol
4	19.4	88.2	108930	9	AL513122	Human DNA
5	19.4	88.2	193050	1	AL646062	Ralstonia
6	19.4	88.2	193168	2	AC011969	Homo sapi
7	19.4	88.2	207392	2	AC006558	Homo sapi
8	19	86.4	15982	1	AE004779	Pseudomon
9	19	86.4	168862	9	AC073464	Homo sapi
10	19	86.4	170237	9	AL137070	Human DNA
11	18.8	85.5	324	6	AX315358	Sequence
12	18.8	85.5	2540	10	BC011256	Mus muscu
13	18.8	85.5	25010	9	AL353092	Human DNA
14	18.8	85.5	176253	2	AP001078	Homo sapi
15	18.8	85.5	193487	2	AP001793	Homo sapi
16	18.8	85.5	210734	2	AC015676	Homo sapi
17	18.8	85.5	212055	2	AP000899	Homo sapi
18	18.4	83.6	4091	1	AF323675	Bradyrhiz
19	18.4	83.6	5162	1	AF151965	Pseudomon
20	18.4	83.6	8660	1	AF153207	Pseudomon
21	17.8	80.9	328	4	CFAQP2X1	Y10638 C.familiari
22	17.8	80.9	496	6	AX015318	Sequence
23	17.8	80.9	834	6	AX166558	Sequence
24	17.8	80.9	1140	10	MFEATP2	AF023257 Mus muscu
25	17.8	80.9	1590	6	E01204	DNA sequence
26	17.8	80.9	1938	6	AR168888	Sequence
27	17.8	80.9	1938	6	AX101337	Sequence
28	17.8	80.9	1941	6	AR168887	Sequence
29	17.8	80.9	1941	6	AX015319	Sequence
30	17.8	80.9	1941	6	AX015323	Sequence
31	17.8	80.9	1941	6	AX101336	Sequence
32	17.8	80.9	1941	10	MM015976	U15976 Mus musculu
33	17.8	80.9	2010	9	AK057960	Homo sapi
34	17.8	80.9	2022	9	AK026727	Homo sapi
35	17.8	80.9	2040	9	BC013149	Homo sapi
36	17.8	80.9	2122	9	D83735	Homo sapien
37	17.8	80.9	2149	9	AK024835	Homo sapi
38	17.8	80.9	2222	6	AX015317	Sequence
39	17.8	80.9	2776	6	AX230581	Sequence 1
40	17.8	80.9	2816	6	I84559	Sequence 1
41	17.8	80.9	2818	9	HS081984	Human endot
42	17.8	80.9	3217	6	AR168902	Sequence
43	17.8	80.9	3217	6	AX101372	Sequence
44	17.8	80.9	3694	6	AR168893	Sequence
45	17.8	80.9	3694	6	AX101354	Sequence

ALIGNMENTS

RESULT 1	AR172775	AR172775	Sequence 6 from patent	US 6303345	23 bp	DNA	linear	PAT 17-DEC-2001
LOCUS	AR172775	Sequence 6 from patent	US 6303345	23 bp	DNA	linear	PAT 17-DEC-2001	
DEFINITION	Sequence 6 from patent	US 6303345	23 bp	DNA	linear	PAT 17-DEC-2001		
ACCESSION	AR172775	Sequence 6 from patent	US 6303345	23 bp	DNA	linear	PAT 17-DEC-2001	
VERSION	AR172775.1	GI:17912266						
KEYWORDS	Unknown.							
SOURCE	Unknown.							
ORGANISM	Unknown.							
REFERENCE	1 (bases 1 to 23)							
AUTHORS	Rohde, W., Becker, D., Randles, J.W., Hehn, A. and Salamini, F.							
TITLE	Use of a virus DNA as promoter							
JOURNAL	Patent: US 6303345-A 6 16-OCT-2001;							
FEATURES	Location/Qualifiers							
source	1..23							
BASE COUNT	4 a 8 c 9 g 2 t							
ORIGIN	/organism="unknown"							

Query Match 100.0%; Score 22; DB 6; Length 23;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgcccagccgaagcctggg 22
 Db 1 CTGCCAGGCCGAGGCTGGG 22

RESULT 2
 ARL172770 ARL172770 1291 bp DNA linear PAT 17-DEC-2001
 LOCUS Sequence 1 from patent US 6303345.
 DEFINITION
 ACCESSION ARL172770
 VERSION ARL172770.1 GI:17912261
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 UNCLASSIFIED.
 1 (bases 1 to 1291)
 Rohde, W., Becker, D., Randless, J. W., Hehn, A. and Salamini, F.
 TITLE Use of a virus DNA as promoter
 JOURNAL Patent: US 6303345-A 1 16-OCN-2001;
 FEATURES Location/Qualifiers
 1..1291
 /organism="unknown"

BASE COUNT 336 a 323 c 332 g 300 t
 ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1291;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgcccagccgaagcctggg 22
 Db 1045 CTGCCAGGCCGAGGCTGGG 1066

RESULT 3
 CFDCG 1291 bp ss-DNA circular VRL 02-AUG-1993
 LOCUS Coconut foliar decay virus, complete genome.
 DEFINITION
 ACCESSION M29963
 VERSION M29963.1 GI:323306
 KEYWORDS circular; complete genome.
 SOURCE Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.
 ORGANISM Coconut foliar decay virus
 Viruses; ssDNA viruses; Nanovirus.

REFERENCE 1 (bases 1 to 1291)
 AUTHORS Rohde, W., Randless, J. W., Langridge, P. and Hanold, D.
 TITLE Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus
 JOURNAL Virology 176, 648-651 (1990)
 MEDLINE 90266484
 COMMENT Draft entry and printed sequence for [1] kindly submitted by W. Rohde, 15-MAR-1989, for release after publication.

FEATURES
 source 1..1291
 /organism="Coconut foliar decay virus"
 /db_xref="taxon:12474"
 40..70
 /note="stem-loop structure"
 103..975
 /note="ORF 1"
 /codon_start=1
 /protein_id="AAA42894.1"
 /db_xref="GI:323307"
 /translation="MGSSIRWCFITLVETEEAANVRRITSLNVAIYGVDEVAPS
 TQRLHQLGRIHLTKRGLQGLTVLGNDRHLEPTRGSDQNRDYSKERVYLLHGVV
 TRPVKRPRLAQRFAEPDELRELDPGTRRCVVGASVETWRAENPFPFPHNQ
 LEVLSAIGPADRTILTCWRDGDGKSVFAYLGLKPDWFTYTCGGRKDVLYIIE
 DKRNILIDVPRCNLEYLNALLECVKNRAFSDDKYEPLSLYGLDHFVHLVFNVLDP
 YLKISRDRIKLWNI"
 314..775

CDS

/note="ORF 2"
 /codon_start=1
 /protein_id="AAA42895.1"
 /db_xref="GI:323308"
 /translation="MTGFTWSRPVPTNRIETTVRRNGCFSTESRLVLESKGDHWP
 DLRNLMNSAMKQADTDALYTELRWNGQDPLKIRSHFHTTIGSLKCCCLRSQSRT
 IAOSSGYADETEETGSPCLPNISDSSPTGSHVVEPERTYCTSTRTQNEI"
 complement(422..568)
 /note="ORF 6"
 /codon_start=1
 /protein_id="AAA42896.1"
 /db_xref="GI:323309"
 /translation="MEMGTDORPILSTPPKLRVQIFGIRLGLPGGVHQPQIVGP
 IVAF"
 639..797
 /note="ORF 3"
 /codon_start=1
 /protein_id="AAA42897.1"
 /db_xref="GI:323310"
 /translation="MRTRRRRREVRCQISRTQARLYLHMWNQKGRIVPVHGRPKTK
 FNPRTQV"
 complement(823..987)
 /note="ORF 5"
 /codon_start=1
 /protein_id="AAA42898.1"
 /db_xref="GI:323311"
 /translation="MTHTLNIFQYFVSPADFOIIRQDICKYEHMHVPEPKITKGFVFF
 RTECPVINTF"
 1098..1286
 /note="ORF 4"
 /codon_start=1
 /protein_id="AAA42899.1"
 /db_xref="GI:323312"
 /translation="MNRVMGGPTIKDSIWIRINLLCLOCTOPLSTSPSIQVSSLLEKK
 AASLYLPSICFCAIGRLS"

CDS

CDS

CDS

CDS

BASE COUNT 336 a 323 c 332 g 300 t
 ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 1291;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgcccagccgaagcctggg 22
 Db 1045 CTGCCAGGCCGAGGCTGGG 1066

RESULT 4
 AL513122/c 108930 bp DNA linear PRI 26-JUN-2001
 LOCUS Human DNA sequence from clone RP11-477J21 on chromosome 9, complete
 DEFINITION sequence.

ACCESSION AL513122
 VERSION AL513122.7 GI:14572568
 KEYWORDS HTG.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 108930)
 AUTHORS Laird, G.

TITLE Direct Submission
 JOURNAL Submitted (26-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk

COMMENT On Jun 27, 2001 this sequence version replaced gi:14270608.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; SW:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep>. This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>. RP11-477J21 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-477J21. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP11-477J21 is at 108930 in this sequence. The true left end of clone RP11-342H3 is at 29985 in this sequence. The true right end of clone RP11-165P4 is at 100 in this sequence.

FEATURES	Location/Qualifiers
source	1..108930
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="9"
	/clone="RP11-477J21"
	/clone_lib="RP11-11.2"
repeat_region	1..187
repeat_region	/note="MPC21B repeat: matches 6..201 of consensus" 605..725
repeat_region	/note="L2 repeat: matches 2624..2746 of consensus" 879..1178
repeat_region	/note="AluJb repeat: matches 1..309 of consensus" 1187..1672
repeat_region	/note="L1MC4 repeat: matches 7492..7975 of consensus" 1673..1972
repeat_region	/note="AluX repeat: matches 1..304 of consensus" 1973..2675
repeat_region	/note="L1MC4 repeat: matches 6783..7492 of consensus" 3053..3202
repeat_region	/note="MIR repeat: matches 82..248 of consensus" 4252..4523
repeat_region	/note="AluSg repeat: matches 1..283 of consensus" 4641..4753
repeat_region	/note="L2 repeat: matches 2589..2710 of consensus" 5008..5134
repeat_region	/note="L1MB1 repeat: matches 6042..6168 of consensus" 5296..5413
repeat_region	/note="HERVL repeat: matches 4908..5022 of consensus" 5414..5637
repeat_region	/note="MLTIG repeat: matches 1..216 of consensus" 5638..5940
repeat_region	/note="AluSg repeat: matches 1..302 of consensus" 5941..6209
repeat_region	/note="MLTIG repeat: matches 216..510 of consensus" 6210..6327
repeat_region	/note="HERVL repeat: matches 4794..4908 of consensus" 6328..6781
repeat_region	/note="MLTJ repeat: matches 1..516 of consensus" 6782..7153
repeat_region	/note="HERVL repeat: matches 4402..4794 of consensus" 7157..7314
repeat_region	/note="FRAM repeat: matches 1..158 of consensus" 8178..8363
repeat_region	/note="MLTJ repeat: matches 1..200 of consensus" 8396..8634
repeat_region	/note="MLTIC repeat: matches 220..466 of consensus" 8664..8966
repeat_region	/note="AluSg repeat: matches 1..308 of consensus" 9734..10026
repeat_region	/note="AluJb repeat: matches 1..304 of consensus" 10611..11123
repeat_region	/note="MLT2B repeat: matches 1..448 of consensus" 12545..12786
repeat_region	/note="MLTIC repeat: matches 1..229 of consensus" 13425..13855
repeat_region	/note="L1 repeat: matches 4205..4658 of consensus" 13905..14075
repeat_region	/note="MIR repeat: matches 50..216 of consensus" 14560..14653
repeat_region	/note="MIR repeat: matches 152..250 of consensus" 15446..15938
repeat_region	/note="MLTID repeat: matches 1..502 of consensus" 16000..16056
repeat_region	/note="L1MB1 repeat: matches 6105..6162 of consensus" 16057..16331
repeat_region	/note="AluSg repeat: matches 1..275 of consensus" 16332..17014
repeat_region	/note="L1MB1 repeat: matches 5442..6105 of consensus" 17094..17197
repeat_region	/note="L1MC/D repeat: matches 5704..5810 of consensus" 17803..18108
repeat_region	/note="AluJo repeat: matches 7..308 of consensus" 18212..18370
repeat_region	/note="FAM repeat: matches 1..159 of consensus" 18400..18716
repeat_region	/note="L1MC5 repeat: matches 7610..7925 of consensus" 20637..20795
repeat_region	/note="AluSg/x repeat: matches 140..291 of consensus" 20939..21587
repeat_region	/note="L2 repeat: matches 1949..2748 of consensus" 21631..21947
repeat_region	/note="AluJb repeat: matches 1..299 of consensus" 22652..22956
repeat_region	/note="AluSg repeat: matches 1..305 of consensus" 23007..23286
repeat_region	/note="AluSg repeat: matches 5..283 of consensus" 23327..23364
repeat_region	/note="19 copies 2 mer gt 100% conserved" 23383..23579
repeat_region	/note="AluJo repeat: matches 1..193 of consensus" 23639..23950
repeat_region	/note="AluSg repeat: matches 1..298 of consensus" 24131..24437
repeat_region	/note="AluSg repeat: matches 1..305 of consensus" 24478..24527
repeat_region	/note="L2 repeat: matches 2698..2747 of consensus" 24643..24880
repeat_region	/note="L2 repeat: matches 2248..2476 of consensus" 24954..25299
repeat_region	/note="MLT1 repeat: matches 68..410 of consensus" 25300..25344
repeat_region	/note="L2 repeat: matches 2197..2235 of consensus" 25345..25707
repeat_region	/note="MLT1A1 repeat: matches 1..365 of consensus" 25708..25752
repeat_region	/note="L2 repeat: matches 2151..2197 of consensus" 27394..27701
repeat_region	/note="AluJ repeat: matches 1..309 of consensus" 27703..27989
repeat_region	/note="AluJb repeat: matches 1..295 of consensus" 28105..28317
repeat_region	/note="L1MC/D repeat: matches 5617..5822 of consensus" 28758..29049
repeat_region	/note="L1MA4 repeat: matches 5769..6074 of consensus" 29050..29372
repeat_region	/note="AluSg repeat: matches 1..311 of consensus" 29373..29603
repeat_region	/note="L1MA4 repeat: matches 6074..6300 of consensus" 29604..29934

```

repeat_region 29698..29949
/note="MIR repeat: matches 12..262 of consensus"
repeat_region 29964..30044
/note="MIR repeat: matches 127..197 of consensus"
repeat_region 30045..30356
/note="AluX repeat: matches 3..312 of consensus"
repeat_region 30357..30390
/note="MIR repeat: matches 92..127 of consensus"
repeat_region 30668..30932
/note="AluX repeat: matches 40..298 of consensus"
repeat_region 31360..31519
/note="L2 repeat: matches 2540..2710 of consensus"
repeat_region 31530..31778
/note="L2 repeat: matches 2483..2746 of consensus"
repeat_region 31834..32101
/note="L2 repeat: matches 2243..2508 of consensus"
repeat_region 33093..33156
/note="MIR repeat: matches 86..151 of consensus"
repeat_region 33878..34052
/note="MIR repeat: matches 20..199 of consensus"
repeat_region 34281..34362
/note="MIR repeat: matches 98..194 of consensus"
repeat_region 34707..34926
/note="MIR repeat: matches 20..262 of consensus"
repeat_region 35373..35551
/note="L2 repeat: matches 2520..2710 of consensus"
repeat_region 35560..35847
/note="MR63B repeat: matches 1..278 of consensus"
repeat_region 35848..36062
/note="THEIC repeat: matches 166..371 of consensus"
repeat_region 36063..36363

Query Match 88.28; Score 19.4; DB 9; Length 108930;
Best Local Similarity 95.28; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ctgccagcggccgaaggctg 21
||||||| |||||
Db 68466 CTGCCAGGCCAAAGGCTGG 68446

RESULT 5
AL646062/c
LOCUS
DEFINITION Ralstonia solanacearum GMI1000 chromosome, complete sequence;
segment 6/19
ACCESSION AL646062
VERSION AL646062.1 GI:17427974
KEYWORDS
SOURCE Ralstonia solanacearum.
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
1 (bases 1 to 193050)
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
Sigquier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 193050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,

```

```

F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
INRA, BP27, 31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES
Location/Qualifiers
1..193050
/organism="Ralstonia solanacearum"
/strain="GMI1000"
/db_xref="taxon:305"
505..771
/gene="RSc0962"
/note="RS04402"
505..771
/gene="RSc0962"
/function="elements of external origin; phage-related
functions and prophages"
/note="product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE BACTERIOPHAGE DNA-BINDING TRANSCRIPTION
REGULATOR PROTEIN"
/protein_id="CAD14664.1"
/db_xref="GI:17427975"
/translation="MKYORAHVAACSQFTPAPVSEATAERFLRLPEVMSACGLS
RSSYGVKSGCFEPVRLTPSSVAWVDSEIQWIAAARQS"
789..1517
/gene="RSc0963"
/note="RS04401"
789..1517
/gene="RSc0963"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD14665.1"
/db_xref="GI:17427976"
/translation="MKONTVAALPENQAALTRHQARGKFEARDEADRIADMOK
HRTAADASETAQQAQAEAKLMRSTKTPQDLRLDKAKERAAYATAEDYRAIYAEFE
EAHVDAKITAGVAKLDENAAVTVLTQYADTLMSEAKOMLVPLLRATIAQVRAISGA
TFDNAWEYMNESAAALARMYGVIKQAFKFDRAQDAVLQAAIRPADLDRFIFP
SPGATHVNTLRKAAQOQGRGA"
1811..2032
/gene="RSc0964"
/note="RS04403"
1811..2032
/gene="RSc0964"
/function="elements of external origin; phage-related
functions and prophages"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE BACTERIOPHAGE TRANSCRIPTIONAL
ACTIVATOR-RELATED TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD14666.1"
/db_xref="GI:17427977"
/translation="MNCPRGSGVAIRSRPVSRTIRELYCQSNVLCGHTFVSLVEV
VRLTSPSTPDPEVAHQLAGRSEPVAAHV"
2400..2576

```

```

/ gene="RSC0965"
/ note="RS04404"
2400. .2576
/ gene="RSC0965"
/ function="miscellaneous; unknown"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="HYPOTHETICAL PROTEIN"
/ protein_id="CAD14667.1"
/ db_xref="GI:17427978"
/ translation="MLDAPARPEQPAFPQILAIVRTALRDVAAPTRDASLDVAGAAAL
LAVAAIAQAARRRG"
2569. .2892
/ gene="RSC0966"
/ note="RS04400"
2569. .2892
/ gene="RSC0966"
/ function="miscellaneous; hypothetical/global homology"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Framed"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="CONSERVED HYPOTHETICAL PROTEIN"
/ protein_id="CAD14668.1"
/ db_xref="GI:17427979"
/ translation="MCKRNHPHWMALRQSLTAQLAQQAQPAQTPFAATLARLETVS
VTEML"
2903. .5704
/ gene="RSC0967"
/ note="RS04399"
2903. .5704
/ gene="RSC0967"
/ function="miscellaneous; unknown"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Framed"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="HYPOTHETICAL PROTEIN"
/ protein_id="CAD14669.1"
/ db_xref="GI:17427980"
/ translation="WHTQSPALTFWFDLKRIDLHLAERLLRRSGAKGNVHSPH
HTRASLSIFRGRGWKDWSTEGGSCIDLVQHCMPDVAITPMEAKLLGQWYGIPI
APASALLAKSTETIARAQANPEPAVAVLAGVDEAVSRRAITQAGLGNWNS
PKVAGEAGHPAAFIYRAMDTARVAVDLRYADPAINGNVKTCQGDKLGHGWT
DARRLHAHTYIVFSPINALSVECCHPGTAALFALGTANVDKIDWTFURGRVVI
ALDHTDPVNERMTPGLAAWKLSEALTAADIGAMLVDMODWEEGEDINDVLOAHG
AELTARMRLLEAWLIPGMPGGERLSCTRVLPDHPDGIYWRFRVEDFTQYVEKFR
HGDATQREVGVNDKLVLEWGRKFHIMWSPQFARMVNVLEAADLGARDVNVFVL
AWRGELAALEGSCDYFEPQKCLYNNRFRSTQASARAVIDAYQATFKGNAATA
LWALGLTKTVLGYPHLOMAKAGKSKLLESLQATSFQVLSGOMLKTDRHRA
VSVYTSHPGVWDFEKLPAVLSEIDGLQSTYRFTRVCAALTPLYLMCAPVLLAGE
EVDVFSLOSKTCRTSLSDKOGAILPRNLPPFPVWALOFLASQOPERIDLHAFVD
VLSLSRAETNDATARNVENVAAIMTAWALLAEFAQIDVEGQFIDLLVEMNHLTA
ETDGRLPFWIMLELLSELDAGRFHPHCWDTTQDSGREMALFLRPSVMDHLSTA
MLRQKFDALPKTARVEFKQLQSGVVPQHGGRKLEDEKRLRGRCAHMAAISLPR
/ complement(6298. .7473)
/ gene="RSC0968"
/ note="RS04398"
2903. .5704
/ gene="RSC0968"
/ function="miscellaneous; unknown"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="HYPOTHETICAL PROTEIN"
/ protein_id="CAD14670.1"
/ db_xref="GI:17427981"
/ translation="MPLTDLRSARNAKPOKPYRLADGLGMYLEVMPNGSKYWLKRYRI
AGKERRALGVPSVLAQARTARDARRLIAAGTDPQOARKEAKLATTTARTTFFEV
VARWEKOSATWPTYATKVLHSLTETDFPDLGARPTETAPOMLATLKVSRDA
LETVARLOQCSAVFRVAVTGRAHNCTDLKGAFKAKVQHVARVOANVEPTLIK
IRGYDGLITRIALQFMALTVFRTKELGAQWSEIDFETEWRIPAEKMKRPHFVP
LSWQIALLRARAAKMTGRQVYFQSPGRGTHISNNTILYALRYLGHSMRTGHGFRGL
ASTMLHEQGFNPDIEROLAHARNKVRAYNHAQYLPERRRMQHWADILDQLAQD"
complement(7649. .7723)
/ gene="tRNA-argW"
tRNA
Query Match 88.2%; Score 19.4; DB 1; Length 193050;
Best Local Similarity 95.2%; Pred. No. 3.le+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 tgccccagccgaagcctggg 22
|||||
Db 171301 TGCCAGCGCGAGGCTCGG 171281
|||||
RESULT 6
AC011969/c
LOCUS
DEFINITION
Homo sapiens chromosome 4 clone RP11-520J8 map 4, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC011969
AC011969.3 GI:7630668
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 193168)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 4, clone RP11-520J8
Unpublished
2 (bases 1 to 193168)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgaiter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Gallagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,K., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
COMMENT
Submitted (17-OCT-1999), Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 21, 2000 this sequence version replaced gi:7008795.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

```


* be preserved.
 1 3313: contig of 3313 bp in length
 * gap of unknown length
 * 3314 9632: contig of 6319 bp in length
 * gap of unknown length
 * 9633 24168: contig of 14536 bp in length
 * gap of unknown length
 * 24169 38366: contig of 14198 bp in length
 * gap of unknown length
 * 38367 56172: contig of 17806 bp in length
 * gap of unknown length
 * 56173 87103: contig of 30931 bp in length
 * gap of unknown length
 * 87104 207392: contig of 120289 bp in length.
 Location/Qualifiers
 1. 207392
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /map="g"
 /clone="hRPK 477 J-21"
 /clone_lib="PC1-11 human BAC library"
 BASE COUNT 57991 a 44607 c 45345 g 58166 t 1283 others
 ORIGIN

Query Match 88.2%; Score 19.4; DB 2; Length 207392;
 Best Local Similarity 95.2%; Pred. No. 3e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ctgcccagccgaaggcctgg 21
 |||||
 Db 70048 CTGCCAGCCAAAGGCGCTGG 70028

RESULT 8
 AE004779 15982 bp DNA linear BCT 30-AUG-2000
 LOCUS Pseudomonas aeruginosa PA01, section 340 of the complete
 DEFINITION genome.
 ACCESSION AE004779 AE004091
 VERSION AE004779.1 GI:9949735
 KEYWORDS
 SOURCE Pseudomonas aeruginosa.
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 REFERENCE 1 (bases 1 to 15982)
 AUTHORS Stover,C.K., Pham,X.O., Erwin,A.L., Mizoguchi,S.D., Warriner,P.,
 Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
 Garber,R.L., Golltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
 Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
 Smith,K., Spencer,D., Wong,G.K., Wu,Z., and Paulsen,I.T.
 Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen
 Nature 406 (6799), 959-964 (2000)
 2043737
 2 (bases 1 to 15982)
 REFERENCE Stover,C.K., Pham,X.O., Erwin,A.L., Mizoguchi,S.D., Warriner,P.,
 Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J.,
 Lagrou,M., Garber,R.L., Golltry,L., Tolentino,E.,
 Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
 Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
 Hancock,R.E.W., Wu,Z., Paulsen,I.T., Reizer,J., Sailer,M.H.,
 Wong,G.K.-S., and Olson,M.V.
 Direct Submission
 Submitted (16-MAY-2000) Department of Medicine and Genetics,
 University of Washington Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 Location/Qualifiers
 1. 15982
 /organism="Pseudomonas aeruginosa"
 /strain="PA01"

gene
 /db_xref="taxon:287"
 163. 1002
 /gene="glpF"
 /note="PA3581"
 163. 1002
 /gene="glpF"
 /codon_start=1
 /transl_table=11
 /product="glycerol uptake facilitator protein"
 /protein_id="AAG06969.1"
 /db_xref="GI:9949736"
 /translation="MTTAAPTPSLFGQCIAEFLGTALLIFFGTGCAALKVAGASFG
 WEISLWGVGSMALYLSAGVSGAHLNPAVSLALWLFAGFEGRLKLPFYITAAQVAF
 AALVYTLVSSLFTEFEQANIVRGSODSLASVSTYFPHALSVGQAFIVVVITA
 ILMVIMALTDDGNLPRGLAPLILGLLIIVIGSAMGLTGFAMNPARDFGPKLMTY
 LAGWPIAFTGTGREIPYFLVPIFAPILGACLGAGGYRVLIARHLPSAAAPAEPEKV
 RAS"
 1042. 2559
 /gene="glpK"
 /note="PA3582"
 1042. 2559
 /gene="glpK"
 /codon_start=1
 /transl_table=11
 /product="glycerol kinase"
 /protein_id="AAG06970.1"
 /db_xref="GI:9949737"
 /translation="MTDKHKYVVALDQGTSSRAIVFDRDANVWSQAOREPAQFYP
 QAGVHDHPMEIWAQTOSTLVEALQAQSIHEDVAAIGITNORETTVMWDRSGRPIH
 NAIWQCRRAAICAOLKRGLEDY IRETTGLVTDYFSGTKLWILDNVEGARERAR
 NGDLLEGTIDFWLWKLTEKGVHVDYTNASRTWLFNHSRDWDMLEVDLIPRSL
 PEVNSSEVIGNARIGGVGGELPIAGIQDQAAALFGQMCVPEQAKNTYGTGCFLL
 MHTGKAVKSTHGLTTIACGPRGEVYALEGAVNGSGTQWMLDKLYINDSFDE
 YFATKVKDSNGVLYPAFTGLGAPYWPYARGAVFLTRGVKADHLIRATLESAYOT
 RDVLDMQDAGERLRALRVDDGAVANFLMFOADILGTFRVPRVRETATGAAVL
 AGLACFWSSLDLKSXAVIERVPECEDEPRREKLYAGMKKAVERTRGWDDGEL"
 2765. 3520
 /gene="glpR"
 /note="PA3583"
 2765. 3520
 /gene="glpR"
 /codon_start=1
 /transl_table=11
 /product="glycerol-3-phosphate regulon repressor"
 /protein_id="AAG06971.1"
 /db_xref="GI:9949738"
 /translation="MNLPPRQQSIIIEVRERGYLSIEEQAQMEAVTPOTLRDINOLA
 EQGLIRRYHGGAAVDSSIENTAVTMRADOMRDEKQRTAEAVASLVPDNASLFTNIGTT
 TEALARALLNHRNLIITNHLVAAATLSAKEDFEVLVAGGTVRSDDGIVGQAAVDFTQ
 QFVDFALVIGSIDEDGSLDDEFYQEVRYSOAILDNAROVFLAADSCKFGRNAVRL
 GPTSLNVRVTSDSPPPAAIARLMAQHKVHLDLV"
 3798. 5336
 /gene="glpD"
 /note="PA3584"
 3798. 5336
 /gene="glpD"
 /codon_start=1
 /transl_table=11
 /product="glycerol-3-phosphate dehydrogenase"
 /protein_id="AAG06972.1"
 /db_xref="GI:9949739"
 /translation="MSQAHTPSAPLAEVVDVAVVGGINGVGTAAADAAGRLSVFLCE
 QHDLAQTSSAKSLIHGRLYLEHFEFLVRELAERVELLAKAPHIVKPKFVLPH
 RPHLRPAWIRAGLFLYDLHLGKREKLPAKSLRGTSSPLKAEIRGFEYSQCAVDDA
 RLVLVNAISAREHGAHVHTRCVRARRSKGLWHLHLERSDDGSLYIRARALVNAAG
 WVARFTQDGLKQSPYGIIRLIQSGHIVPKLYEGEHAYILQNDREIRIVPAIYLDRT
 MICTDREYQDPAKVAIBEEFTAYLLQVNAHFKQOLAADTLHSFAGVRLCDDES
 DEPSATITRYTLLSAGNEPPLLVFGKLTTRYKLAESALTQLOFPFANLGPAAWTA
 KAPLPGEOQMSVEALTEQLANRYAMLDRELALRWARTYGTWRWLLDGVNAGDE
 HIGGLYAREVDYLCKHEWAQDAEDILWRRSKLGLEFLSPSQOVRUGLYQSEBHPHRP
 VHAA"
 5469. 5798
 /gene="glpM"


```
repeat_region 29506..29537 /rpt_family="AT_rich"
repeat_region 29984..30783 /rpt_family="L1"
repeat_region 30871..31273 /rpt_family="L1"
repeat_region 31261..31285 /rpt_family="AT_rich"
repeat_region 31807..31839 /rpt_family="AT_rich"
repeat_region 32015..32038

Query Match 86.4% Score 19; DB 9; Length 168862;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 gccacggccgaagcgctgg 21
Db 65377 GCCCAGGCCGAGGCGCTGG 65395

RESULT 10
AL137070/c
LOCUS
DEFINITION Human DNA sequence from clone Rp11-251017 on chromosome 9 Contains
1 calponin 2 (CNN2) pseudogenes, a gene for a novel protein similar
to aquaporin 7 (AQ7), part of gene for a novel protein similar to
methyltetrahydrofolate dehydrogenase (NADP+ dependent), folate
methyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate
synthetase (MTHFD1) and a CpG island, complete sequence.
AL137070
AL137070.9 GI:9367917
HTG; calponin; CNN2; CpG island.
human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 170237)
Skuce.C.
Direct Submission
Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humqueres@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Jul 22, 2000 this sequence version replaced gi:8977609.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit subsequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
This sequence is the entire insert of clone Rp11-251017 This
sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. Rp11-251017 is from
the library RPCI-11.1 constructed by the group of Pieter de Jong.
For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6.
Location/Qualifiers
1..170237

FEATURES
source
misc_feature 3..433 /note="match: GSS: Em:AQ313358"
misc_feature 5..564 /note="match: GSS: Em:AQ884352"
misc_feature 7..785 /note="match: GSS: Em:AQ741764"
misc_feature 7..439 /note="match: GSS: Em:AQ182496"
repeat_region 92..384 /note="AluJo repeat: matches 1..288 of consensus"
misc_feature complement(527..1042)
repeat_region 1467..1634 /note="match: GSS: Em:AQ174050"
repeat_region 2090..2613 /note="2 copies 84 mer 81% conserved"
misc_feature complement(2326..3058) /note="L1PB2 repeat: matches 5614..6155 of consensus"
misc_feature complement(2332..3081) /note="match: GSS: Em:AQ078000"
misc_feature complement(2366..3084) /note="match: GSS: Em:AQ749938"
misc_feature complement(2586..3082) /note="match: GSS: Em:AQ530925"
misc_feature complement(2606..3087) /note="match: GSS: Em:AQ435304"
misc_feature complement(2638..3058) /note="match: GSS: Em:AQ416033"
misc_feature complement(2642..3082) /note="match: GSS: Em:AQ057987"
misc_feature complement(2670..3084) /note="match: GSS: Em:AQ200338"
misc_feature complement(2676..3081) /note="match: GSS: Em:AQ494022"
misc_feature complement(2677..3085) /note="match: GSS: Em:AQ719378"
misc_feature 3059..3150 /note="match: GSS: Em:AQ410685"
misc_feature 3068..3150 /note="match: GSS: Em:AQ038020"
misc_feature 3068..3362 /note="match: GSS: Em:AQ239460"
misc_feature 3088..3353 /note="match: GSS: Em:AQ111808"
misc_feature 3089..3264 /note="match: GSS: Em:AQ810980"
misc_feature 3101..3363 /note="match: GSS: Em:AQ831431"
misc_feature 3101..3479 /note="match: GSS: Em:AQ559024"
misc_feature 3106..3531 /note="match: GSS: Em:AQ493601"
misc_feature 3115..3699 /note="match: GSS: Em:AQ033506"
misc_feature 3155..3616 /note="match: GSS: Em:AQ045861"
misc_feature 3155..3700 /note="match: GSS: Em:AQ239460"
misc_feature 3155..3700 /note="match: GSS: Em:AQ525695"
misc_feature 3155..3388 /note="match: GSS: Em:AQ038020"
misc_feature 3266..3538 /note="match: GSS: Em:AQ831431"
repeat_region 3348..3735 /note="MER70A repeat: matches 166..564 of consensus"
misc_feature 3355..3572 /note="match: GSS: Em:AQ810980"
repeat_region 4107..4340 /note="L1M4 repeat: matches 5352..5586 of consensus"
```

```

repeat_region 4398..4519
/note="MSTB repeat: matches 1..121 of consensus"
repeat_region 4520..4733
/note="MER30 repeat: matches 1..230 of consensus"
misc_feature complement(4614..5113)
/note="match: GSS: Em:AQ547317"
misc_feature complement(4616..5130)
/note="match: GSS: Em:AQ380868"
misc_feature complement(4616..5127)
/note="match: GSS: Em:AQ748099"
misc_feature complement(4616..5124)
/note="match: GSS: Em:AQ682825"
misc_feature complement(4661..5373)
/note="match: GSS: Em:AQ194341"
misc_feature 4734..5306
/note="match: GSS: Em:AQ269571"
repeat_region 4734..5035
/note="MSTB repeat: matches 121..426 of consensus"
misc_feature complement(4775..5372)
/note="match: GSS: Em:AQ418853"
misc_feature complement(4835..5359)
/note="match: GSS: Em:AQ664754"
misc_feature complement(4837..5369)
/note="match: GSS: Em:AQ454594"
misc_feature complement(4860..5375)
/note="match: GSS: Em:AQ346755"
misc_feature complement(4865..5375)
/note="match: GSS: Em:AQ549101"
misc_feature complement(4885..5367)
/note="match: GSS: Em:AQ693755"
misc_feature 5131..5586
/note="match: GSS: Em:AQ409721"
misc_feature 5131..5574
/note="match: GSS: Em:AQ798688"
misc_feature 5131..5633
/note="match: GSS: Em:AQ693759"
misc_feature 5170..5513
/note="match: GSS: Em:AQ282181"
misc_feature 5381..5493
/note="match: GSS: Em:AQ808738"
misc_feature 5386..5658
/note="match: GSS: Em:AQ426687"
misc_feature 5389..5789
/note="match: GSS: Em:AQ520858"
misc_feature 5393..5859
/note="match: GSS: Em:AQ180925"
misc_feature 5404..5659
/note="match: GSS: Em:AQ764635"
misc_feature 5431..5593
/note="match: GSS: Em:B75305"
misc_feature 5450..5792
/note="match: GSS: Em:B82392"
misc_feature 5497..5900
/note="match: GSS: Em:AQ808738"
repeat_region 5775..6068
/note="AluJo repeat: matches 1..282 of consensus"
misc_feature 5793..5946
/note="match: GSS: Em:AQ520858"
misc_feature complement(6293..6887)
/note="match: GSS: Em:B65313"
misc_feature complement(6419..6895)
/note="match: GSS: Em:AQ232792"
misc_feature complement(6420..6876)
/note="match: GSS: Em:AQ178179"
misc_feature complement(join(6671..7023,7998)..8098))
/note="match: GSS: Em:AQ278128"
misc_feature complement(join(6703..7023,7998)..8166))
/note="match: GSS: Em:AQ571063"
misc_feature complement(6772..7032)
/note="match: GSS: Em:AQ512891"
misc_feature 6875..7321
/note="match: GSS: Em:AQ759250"
misc_feature 6902..7238

```

```

misc_feature /note="match: GSS: Em:AQ770277"
6987..7346
repeat_region /note="match: GSS: Em:AQ228275"
7030..7997

Query Match 86.4%; Score 19; DB 9; Length 170237;
Best Local Similarity 100.0%; Pred. NO. 4.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gccagagccgaagcctgg 21
|||||
Db 26857 GCCAGGCCGAGGCCTGG 26839
|||||

RESULT 11
AX315358/c
LOCUS AX315358 324 bp DNA linear PAT 14-DEC-2001
DEFINITION Sequence 8343 from Patent WO0190366.
ACCESSION AX315358
VERSION AX315358.1 GI:17898914
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Leach,M.D. and Shinkets,R.A.
TITLE Human polynucleotides and polypeptides encoded thereby
JOURNAL Patent: WO 0190366-A 8343 29-NOV-2001;
Curagen Corporation (US)
FEATURES
source 1..324
Location/Qualifiers
BASE COUNT 82 a 84 c 89 g 69 t
ORIGIN
/note="match: GSS: Em:AQ770277"
6987..7346
repeat_region /note="match: GSS: Em:AQ228275"
7030..7997

Query Match 85.5%; Score 18.8; DB 6; Length 324;
Best Local Similarity 90.9%; Pred. NO. 3.8e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccagccgaagcctggg 22
|||||
Db 71 CTGCCAGGCCGAGGCCTGGG 50
|||||

RESULT 12
BC011256/c
LOCUS BC011256 2540 bp mRNA linear ROD 30-JUL-2001
DEFINITION Mus musculus, Similar to plasminogen activator, tissue, clone
MGC:18508 IMAGE:4038404, mRNA, complete cds.
ACCESSION BC011256
VERSION BC011256.1 GI:15030027
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2540)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT NTH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: crapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome

```

Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 24 Row: d Column: 24
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 202109.

FEATURES
source
1. .2540
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:18508 IMAGE:4038404"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
/clone_lib="NCI_CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
108. 1787
/codon_start=1
/product="Similar to plasminogen activator, tissue"
/protein_id="AAH11256.1"
/db_xref="GI:15030028"
/translation="MKRELICVLLGLGLAPLPDQGIHGRFRGRARSYRATCRDEPTQ
TTVQHOSLWRLPMLRSRYEYRCNSGLVQCHSVPRSCSEPCFNGTCQALYFSD
FVQCQDPGFKGKCDIDTRATCEQGIYRTWTSAESGAEICINWNSVLKYPNA
RRNALIKGNHNYCRNDRDLKPMCYVFKAGKYTEFCSTPACPKSGEDCYGKG
VYTGHSILTSOASCLPWSIVLMKSTANTWNTSQALGHRHNYCRNPDGADAPWC
HVMDKRLTWEYCDMSPCSTGRLQYRPFRIKGLYDITSHPMQALFVANKRSP
GERFLCGVLISSCWLSAHCFLEFPNHLKVLGRTYRVVPEEGTFFELKTIY
HEFDDTDYNDIALQLRSQKQCAESSVGTACLPDNLQPLDTECELSGYGKH
EASPFSDRLKAHRLXPSSRCTSOHLFNKVTNNMLCAGDTRSGNQDLHDACOG
DSGGPLVCMINKMTLTGIIISWLGQSGQKQDVPYTKVTNYLWDIHNKQ"

BASE COUNT 699 a 619 c 644 g 578 t

ORIGIN
Query Match 85.5%; Score 18.8; DB 10; Length 2540;
Best Local Similarity 90.9%; Pred. No. 2e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccagcgcgaagcctggg 22
|||||
DB 889 CTGCCAGGCGAGTGCCTGGG 868

RESULT 13
AL353092/c
LOCUS
DEFINITION
Human DNA sequence from clone RPI-180I13 on chromosome 20 Contains
5' end of the HCK gene for hemopoietic cell kinase (protein
tyrosine kinase), contains ESTs, STSs, GSSs and a CpG island,
complete sequence.

ACCESSION AL353092
VERSION AL353092.6 GI:9650539
KEYWORDS HTG; CpG island; HCK; tyrosine kinase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 25010)
Almeida, J.
Direct Submission
Submitted (08-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 1, 2000 this sequence version replaced gi:9187765.
During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr20>
IMPORTANT: This sequence is not the entire insert of clone RPI-180I13 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RFS-836N17 is at 24911 in this sequence. The true right end of clone RPI-3100I3 is at 100 in this sequence. This sequence was finished as follows otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RPI-180I13 is from the library RPI-1 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2.

FEATURES
source
Location/Qualifiers
1. .25010
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RPI-180I13"
/clone_lib="RPI-1"
1. .43
/note="AluJo/FLAM repeat: matches 76. .118 of consensus"
complement(39. .567)
/note="match: GSS: Em:AQ386884"
58. 175
/note="S5 repeat: matches 1. .119 of consensus"
285. .435
/note="MER20 repeat: matches 26. .218 of consensus"
510. .586
/note="L2 repeat: matches 1966. .2045 of consensus"
692. .906
/note="match: GSS: Em:AQ486790"
772. .843
/note="2 copies 36 mer 93% conserved"
1576. .2453
/note="CpG island"
/evidence="not_experimental"
join(1846. .1956,21132. .21252,22789. .22831,23186. .23288, 24093. .>24191)
/gene="HCK"
/note="match: cDNAs: Em:M16591 Em:X62345 Em:S74141
Em:J03023 Em:M16592 Em:M83666 Em:Y00487 Em:X60380
Em:J03579 Em:X67786 Em:AF000300 Em:AF000301 Em:AF000302
Em:M17031 Em:X52822 Em:X57018 Em:X15345 Em:M27454
Em:M57290 Em:X57684 Em:M85043 Em:M19722
match: ESTs: Em:AW307786"
/product="dl180I13.1 (hemopoietic cell kinase)"
/evidence="not_experimental"
1846. .24191
/gene="HCK"
complement(2102. .2519)
/note="match: GSS: Em:AQ382576"
complement(2108. .2557)
/note="match: GSS: Em:AQ461389"
complement(2196. .2570)
/note="match: GSS: Em:AQ240986"

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: <http://hdp.gsc.riken.go.jp/>

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: Rp11-861B19

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 155758 bases at least Q40

Consensus quality: 165815 bases at least Q30

Consensus quality: 170532 bases at least Q20

Insert size: 173753; sum-of-contigs

Quality coverage: 4.13x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1      25481 contig of 25481 bp in length
25582  44141 contig of 18560 bp in length
44242  59207 contig of 14966 bp in length
59308  73163 contig of 13856 bp in length
73264  83431 contig of 10168 bp in length
83532  90995 contig of 7464 bp in length
91096  99746 contig of 8651 bp in length
99847  105997 contig of 6151 bp in length
106098 112950 contig of 5257 bp in length
113051 118074 contig of 5024 bp in length
118175 124012 contig of 5641 bp in length
124113 129369 contig of 5936 bp in length
129470 135110 contig of 5364 bp in length
135211 141146 contig of 5364 bp in length
141247 146610 contig of 5364 bp in length
146711 150822 contig of 4112 bp in length
150923 155230 contig of 4308 bp in length
155331 158760 contig of 3430 bp in length
158861 163143 contig of 2446 bp in length
163244 167768 contig of 2446 bp in length
167869 170314 contig of 1532 bp in length
170415 171946 contig of 1532 bp in length
171947 172046 contig of 100 bp in length
172047 173708 contig of 1662 bp in length
173809 174921 contig of 1113 bp in length
175022 176253 contig of 1232 bp in length

```

Sequence updated (26-May-2000).

NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1      25481: contig of 25481 bp in length
25482 25582: gap of 100 bp
25582 44141: contig of 18560 bp in length
44142 44242: gap of 100 bp
44242 59207: contig of 14966 bp in length
59208 59308: contig of 13856 bp in length
59308 73163: contig of 10168 bp in length
73164 73264: gap of 100 bp
73264 83431: contig of 10168 bp in length
83432 83531: gap of 100 bp
83532 90995: contig of 7464 bp in length
90996 91095: gap of 100 bp
91096 99746: contig of 8651 bp in length
99747 99846: gap of 100 bp

```

```

*      99847 105997: contig of 6151 bp in length
*      105998 106097: gap of 100 bp
*      106098 112950: contig of 5257 bp in length
*      112951 113050: gap of 100 bp
*      113051 118074: contig of 5024 bp in length
*      118075 118174: gap of 100 bp
*      118175 124012: contig of 5641 bp in length
*      124013 124112: gap of 100 bp
*      124113 129369: contig of 5936 bp in length
*      129370 129469: gap of 100 bp
*      129470 135110: contig of 5364 bp in length
*      135111 135210: gap of 100 bp
*      135211 141146: contig of 5364 bp in length
*      141147 141246: gap of 100 bp
*      141247 146610: contig of 5364 bp in length
*      146611 146710: gap of 100 bp
*      146711 150822: contig of 4112 bp in length
*      150823 150922: gap of 100 bp
*      150923 155230: contig of 4308 bp in length
*      155231 155330: gap of 100 bp
*      155331 158760: contig of 3430 bp in length
*      158761 158860: gap of 100 bp
*      158861 163143: contig of 2446 bp in length
*      163144 163243: gap of 100 bp
*      163244 167768: contig of 2446 bp in length
*      167769 167868: gap of 100 bp
*      167869 170314: contig of 1532 bp in length
*      170315 170414: gap of 100 bp
*      170415 171946: contig of 1532 bp in length
*      171947 172046: gap of 100 bp
*      172047 173708: contig of 1662 bp in length
*      173709 173808: gap of 100 bp
*      173809 174921: contig of 1113 bp in length
*      174922 175021: gap of 100 bp
*      175022 176253: contig of 1232 bp in length.

```

FEATURES

source

```

1..176253
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="Rp11-861B19"
1..25481
/note="assembly_fragment"
25582..44141
/note="assembly_fragment"
44242..59207
/note="assembly_fragment"
59308..73163
/note="assembly_fragment"
73264..83431
/note="assembly_fragment"
83532..90995
/note="assembly_fragment"
91096..99746
/note="assembly_fragment"
99847..105997
/note="assembly_fragment"
106098..112950
/note="assembly_fragment clone_end:SP6 vector_side:left"
113051..118074
/note="assembly_fragment"
118175..124012
/note="assembly_fragment"
124113..129369
/note="assembly_fragment"
129470..135110
/note="assembly_fragment"
135211..141146
/note="assembly_fragment"
141247..146610

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```

/note="assembly_fragment"
146711..150822
/note="assembly_fragment"
150923..155230
/note="assembly_fragment"
155331..158760
/note="assembly_fragment"
158861..163143
/note="assembly_fragment"
163244..165114
/note="assembly_fragment"
165215..167768
/note="assembly_fragment"
167869..170314
/note="assembly_fragment"
170415..171946
/note="assembly_fragment"
172047..173708
/note="assembly_fragment"
173809..174921
/note="assembly_fragment"
175022..176253
/note="assembly_fragment"
48857 a 38855 c 39248 g 46787 t 2506 others
ORIGIN

Query Match      85.5%; Score 18.8; DB 2; Length 176253;
Best Local Similarity 90.9%; Pred. No. 5.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctgccaggccgaaggcctggg 22
      ||||| ||||| ||||| |||||
Db 99569 CTGCCAAGGCCAAGGCTGGG 99590

RESULT 15
AP001793
LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-661013 map 18p11.3, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
AP001793
ACCESSION
AP001793.2 GI:8117465
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
Homo sapiens
Homo sapiens DNA, clone:RP11-661013.
ORGANISM
Fukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 193487)
Hattori.M., Ishii.K., Toyoda.A., Taylor.T.D., Hong-Seog.P.,
Fujiyama.A., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.
Homo sapiens 193,487 genomic DNA of 18p11.3
Published Only in DataBase (2000) In press
2 (bases 1 to 193487)
Hattori.M., Ishii.K., Toyoda.A., Taylor.T.D., Hong-Seog.P.,
Fujiyama.A., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.
Direct Submission
Submitted (11-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsr.riken.go.jp,
URL:http://hqp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:7592907.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Web site: http://hqp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-661013
----- Summary Statistics

```

```

Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 176068 bases at least Q40
Consensus quality: 185034 bases at least Q30
Consensus quality: 188618 bases at least Q20
Insert size: 190387; sum-of-contigs
Quality coverage: 4.33x in Q20 bases; sum-of-contigs
-----

```

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 15033 contig of 15033 bp in length
15134 28850 contig of 13717 bp in length
28951 4381 contig of 14431 bp in length
43482 54311 contig of 10830 bp in length
54412 65508 contig of 11097 bp in length
65609 74474 contig of 8866 bp in length
74575 86282 contig of 11708 bp in length
86383 94094 contig of 7712 bp in length
94195 102440 contig of 8246 bp in length
102541 109661 contig of 7121 bp in length
109762 118186 contig of 8425 bp in length
118287 123309 contig of 5023 bp in length
123410 128976 contig of 5567 bp in length
129077 133984 contig of 4908 bp in length
134085 138474 contig of 4390 bp in length
138575 143774 contig of 5200 bp in length
143875 149669 contig of 5795 bp in length
149770 153536 contig of 3767 bp in length
153637 158023 contig of 4387 bp in length
158124 162445 contig of 4322 bp in length
162546 165243 contig of 2698 bp in length
165344 168875 contig of 3532 bp in length
168976 171973 contig of 2998 bp in length
172074 175989 contig of 3916 bp in length
176090 179781 contig of 3692 bp in length
179882 182243 contig of 2362 bp in length
182344 184565 contig of 2222 bp in length
184666 187147 contig of 2482 bp in length
187248 189548 contig of 2301 bp in length
189649 190661 contig of 1013 bp in length
190762 191967 contig of 1206 bp in length
192068 193487 contig of 1420 bp in length
Sequence updated (26-May-2000).

```

* NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 15033: contig of 15033 bp in length
15034 15133: gap of 100 bp
15134 28850: contig of 13717 bp in length
28851 28950: gap of 100 bp
28951 43381: contig of 14431 bp in length
43382 43481: gap of 100 bp
43482 54311: contig of 10830 bp in length
54312 54411: gap of 100 bp
54412 65508: contig of 11097 bp in length
65509 65608: gap of 100 bp
65609 74474: contig of 8866 bp in length
74475 74574: gap of 100 bp
74575 86282: contig of 11708 bp in length
86283 86382: gap of 100 bp
86383 94094: contig of 7712 bp in length
94095 94194: gap of 100 bp

```

* 94195 102440: contig of 8246 bp in length
* 102441 102540: gap of 100 bp
* 109661 109661: contig of 7121 bp in length
* 109662 109761: gap of 100 bp
* 118186 118186: contig of 8425 bp in length
* 118187 118286: gap of 100 bp
* 118287 123309: contig of 5023 bp in length
* 123310 123409: gap of 100 bp
* 123410 128976: contig of 5567 bp in length
* 128977 129076: gap of 100 bp
* 129077 133984: contig of 4908 bp in length
* 133985 134084: gap of 100 bp
* 134085 138474: contig of 4390 bp in length
* 138475 138574: gap of 100 bp
* 138575 143774: contig of 5200 bp in length
* 143775 143874: gap of 100 bp
* 143875 149669: contig of 5795 bp in length
* 149670 149769: gap of 100 bp
* 149770 153536: contig of 3767 bp in length
* 153537 153636: gap of 100 bp
* 153637 158023: contig of 4387 bp in length
* 158024 158123: gap of 100 bp
* 158124 162445: contig of 4322 bp in length
* 162446 162545: gap of 100 bp
* 162546 165243: contig of 2698 bp in length
* 165244 165343: gap of 100 bp
* 165344 168875: contig of 3532 bp in length
* 168876 168975: gap of 100 bp
* 168976 171973: contig of 2998 bp in length
* 171974 172073: gap of 100 bp
* 172074 175989: contig of 3916 bp in length
* 175990 176089: gap of 100 bp
* 176090 179781: contig of 3692 bp in length
* 179782 179881: gap of 100 bp
* 179882 182243: contig of 2362 bp in length
* 182244 182343: gap of 100 bp
* 182344 184565: contig of 2222 bp in length
* 184566 184665: gap of 100 bp
* 184666 187147: contig of 2482 bp in length
* 187148 187247: gap of 100 bp
* 187248 189548: contig of 2301 bp in length
* 189549 189648: gap of 100 bp
* 189649 190661: contig of 1013 bp in length
* 190662 190761: gap of 100 bp
* 190762 191967: contig of 1206 bp in length
* 191968 192067: gap of 100 bp
* 192068 193487: contig of 1420 bp in length.

FEATURES
source

1..193487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-661013"
1..15033
/note="assembly_fragment"
15134..28850
/note="assembly_fragment"
28951..43381
/note="assembly_fragment"
43482..54311
/note="assembly_fragment clone_end:T7 vector_side:left"
54412..65508
/note="assembly_fragment clone_end:SP6 vector_side:left"
65609..74474
/note="assembly_fragment"
74575..86282
/note="assembly_fragment"
86383..94094
/note="assembly_fragment"
94195..102440
/note="assembly_fragment"
102541..109661

misc_feature
109762..118186
/note="assembly_fragment"
misc_feature
118287..123309
/note="assembly_fragment"
misc_feature
123410..128976
/note="assembly_fragment"
misc_feature
129077..133984
/note="assembly_fragment"
misc_feature
134085..138474
/note="assembly_fragment"
misc_feature
138575..143774
/note="assembly_fragment"
misc_feature
143875..149669
/note="assembly_fragment"
misc_feature
149770..153536
/note="assembly_fragment"
misc_feature
153637..158023
/note="assembly_fragment"

Query Match 85.5%; Score 18.8; DB 2; Length 193487;
Best Local Similarity 90.9%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccccagccgaagcctggg 22
||||| ||||| ||||| |||||
Db 186005 CTGCCAAGGCCCAAGGCGCTGGG 186026

Search completed: June 19, 2002, 15:43:14
Job time: 13452 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2002, 11:59:02 : Search time 3798.65 Seconds
(without alignments)
9.944 Million cell updates/sec

Title: US-09-462-955b-1_COPY_655_676

Perfect score: 22
Sequence: 1 ctgcccagccgcaagcctggg 22

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1980.DAT.*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1981.DAT.*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1982.DAT.*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1983.DAT.*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1984.DAT.*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1985.DAT.*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1986.DAT.*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1987.DAT.*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1988.DAT.*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1989.DAT.*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1990.DAT.*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1991.DAT.*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1992.DAT.*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1993.DAT.*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1994.DAT.*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1995.DAT.*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1996.DAT.*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1998.DAT.*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA1999.DAT.*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2000.DAT.*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2001B.DAT.*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	44	20	AA02689
2	22	100.0	44	20	AA02687
3	18.8	85.5	324	21	AA075325
4	18.8	85.5	1592	20	AA027241
5	17.8	80.9	313	22	AA034910
6	17.8	80.9	496	20	AA038123
7	17.8	80.9	524	21	AA049239
8	17.8	80.9	807	23	AA054242
9	17.8	80.9	834	22	AA06749

c 10	17.8	80.9	1092	22	AA01852	Human reproductive cDNA encoding nove
c 11	17.8	80.9	1092	22	AA027197	Lung cancer associ
c 12	17.8	80.9	1658	21	AA018209	Murine FATP1 codin
c 13	17.8	80.9	1938	22	AA089013	Human FATP1 codin
c 14	17.8	80.9	1941	22	AA089012	Human FATP1 codin
c 15	17.8	80.9	2219	20	AA038122	Human FATP protein
c 16	17.8	80.9	2222	20	AA038124	Human FATP variant
c 17	17.8	80.9	2299	22	AA036156	Human musculoskele
c 18	17.8	80.9	2776	22	AA014155	Human EPAS1 DNA us
c 19	17.8	80.9	2816	19	AA000641	cDNA encoding huma
c 20	17.8	80.9	2818	21	AA0294053	DNA encoding human
c 21	17.8	80.9	3217	20	AA000365	Nucleotide sequenc
c 22	17.8	80.9	3217	20	AA089027	Murine FATP1 codin
c 23	17.8	80.9	3694	22	AA000356	Nucleotide sequenc
c 24	17.8	80.9	3694	22	AA089018	Human FATP1 codin
c 25	17.8	80.9	3704	20	AA000352	Nucleotide sequenc
c 26	17.8	80.9	3704	22	AA089010	Human FATP1 codin
c 27	17.8	80.9	6744	20	AA038125	Human FATP genomic
c 28	17.4	79.1	1063	23	AA091080	DNA encoding novel
c 29	17.4	79.1	1116	23	AA085415	DNA encoding novel
c 30	17.4	79.1	1116	23	AA091079	DNA encoding novel
c 31	17.4	79.1	2467	23	ABL12889	Drosophila melanog
c 32	17.4	79.1	2484	23	AA083500	DNA encoding novel
c 33	17.4	79.1	5040	23	ABL12888	Drosophila melanog
c 34	17.2	78.2	401	22	AA096325	Human neuregulin g
c 35	17.2	78.2	401	22	AA097818	Human neuregulin g
c 36	17.2	78.2	493	22	AA075946	Human immune/haema
c 37	17.2	78.2	562	22	ABA61096	Human foetal liver
c 38	17.2	78.2	562	22	ABA28993	Probe #7459 for ge
c 39	17.2	78.2	562	22	AA09389	Human brain expres
c 40	17.2	78.2	562	22	AA032278	Human bone marrow
c 41	17.2	78.2	562	22	AA040995	Probe #9681 used t
c 42	17.2	78.2	993	24	AA019273	Human cDNA for G p
c 43	17.2	78.2	997	22	AA0193426	Human polynucleoti
c 44	17.2	78.2	1851	20	AA027845	Mutant lipase codi
c 45	17.2	78.2	2145	23	AA015727	DNA encoding sulfa

ALIGNMENTS

RESULT 1
AA02689
ID AAX02689 standard; DNA; 44 BP.
XX
AC AAX02689;
XX
DT 10-MAY-1999 (first entry)
XX
CFDV-DNA fragment stem loop repeat region.
DE
DE
KW Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX
OS Coconut foliar decay virus
XX
FH Key Location/Qualifiers
FT stem_loop 4..22
FT /*tag= a
FT /*note= "RPT1"
FT misc_binding 4..10
FT /*tag= b
FT /*note= "Region binds to nucleotides 16 to 22"
FT misc_binding 16..22
FT /*tag= c
FT /*note= "Region binds to nucleotides 4 to 10"
FT stem_loop 27..43
FT /*tag= d
FT /*note= "RPT2"
FT misc_binding 27..32
FT /*tag= e
FT misc_binding 38..43
FT /*note= "Region binds to nucleotides 38 to 43"

FT FT /*tag= f
 PN PN /note= "Region binds to nucleotides 27 to 32"
 XX XX
 PD DE19730502-A1.
 XX 21-JAN-1999.
 XX 16-JUL-1997; 97DE-1030502.
 XX 16-JUL-1997; 97DE-1030502.
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
 PI WPI; 1999-096863/09.
 XX
 XX Coconut foliar decay virus promoters - for gene expression in
 PT bacteria and yeasts
 PT
 XX Disclosure; Fig 4; 14pp; German.
 XX
 XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
 CC that includes the stem-loop structure of CFDV DNA but lacks the
 CC translation start codons of open reading frames ORF1 and/or ORF2. The
 CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
 CC promoter for tissue-specific (especially phloem-specific) gene expression
 CC in plants and for production of chimeric constructs for transient or
 CC stable expression. Certain fragments of CFDV DNA have stronger promoter
 CC activity in E. coli than the CamV 35S promoter.
 XX
 XX Sequence 44 BP; 8 A; 16 C; 17 G; 3 T; 0 other;
 SQ

Query Match 100.0%; Score 22; DB 20; Length 44;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgccaggccgaagcctggg 22
 |||||
 DB 1 ctgccaggccgaagcctggg 22

RESULT 2
 AAX02687
 ID AAX02687 standard; DNA; 44 BP.
 XX AC AAX02687;
 XX DT 10-MAY-1999 (first entry)
 XX DE CFDV-DNA fragment stem loop repeat region.
 XX
 XX Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
 KW yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.
 XX Coconut foliar decay virus

PH Key Location/Qualifiers
 FT stem_loop 4..22
 FT /*tag= a
 FT /note= "RPT1"
 FT misc_binding 4..10
 FT /*tag= b
 FT /note= "Region binds to nucleotides 16 to 22"
 FT misc_binding 16..22
 FT /*tag= c
 FT /note= "Region binds to nucleotides 4 to 10"
 FT stem_loop 27..43
 FT /*tag= d
 FT /note= "RPT2"
 FT misc_binding 27..32
 FT /*tag= e

FT misc_binding /note= "Region binds to nucleotides 38 to 43"
 FT 38..43
 FT /*tag= f
 FT /note= "Region binds to nucleotides 27 to 32"
 XX
 XX DE19730535-A1.
 XX 21-JAN-1999.
 XX 16-JUL-1997; 97DE-1030535.
 XX 16-JUL-1997; 97DE-1030535.
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
 PI WPI; 1999-096867/09.
 XX
 XX Coconut foliar decay virus promoters - for gene expression in
 PT plants, bacteria and yeasts
 PT
 XX Disclosure; Fig 4; 12pp; German.
 XX
 XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
 CC that includes the stem-loop structure of CFDV DNA but lacks the
 CC translation start codons of open reading frames ORF1 and/or ORF2. The
 CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
 CC promoter for tissue-specific (especially phloem-specific) gene expression
 CC in plants and for production of chimeric constructs for transient or
 CC stable expression. Certain fragments of CFDV DNA have stronger promoter
 CC activity in E. coli than the CamV 35S promoter.
 XX
 XX Sequence 44 BP; 8 A; 16 C; 17 G; 3 T; 0 other;
 SQ

Query Match 100.0%; Score 22; DB 20; Length 44;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgccaggccgaagcctggg 22
 |||||
 DB 1 ctgccaggccgaagcctggg 22

RESULT 3
 AAC75325/C
 ID AAC75325 standard; cDNA; 324 BP.
 XX AC AAC75325;
 XX DT 08-FEB-2001 (first entry)
 XX
 XX Human ORFX ORF880 polynucleotide sequence SEQ ID NO:1759.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antiposioratic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; hypothyroidism; SCID; AIDS;
 KW cardiovascular disease; diabetes mellitus; graft vs host disease;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.
 OS
 XX WO200058473-A2.
 PN

05-OCT-2000.
31-MAR-2000; 2000WO-US08621.
31-MAR-1999; 99US-0127607.
02-APR-1999; 99US-0127636.
05-APR-1999; 99US-0127728.
30-MAR-2000; 2000US-0540763.
(CURA-) CURAGEN CORP.
Shimkets RA, Leach M;
WPI; 2000-602362/57.
P-PSDB; AAB41116.
Novel nucleic acids and peptides derived from open reading frame X,
useful for treating e.g. cancers, proliferative disorders,
neurodegenerative disorders and cardiovascular disease -
Claim 5; Page 1386-1387; 5507pp; English.
AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
which represent the human ORFX open reading frames 1 to 3161. The ORFX
sequences have activities such as: cytostatic; hepatotropic; vulnary;
antiparasitic; antiparkinsonian; nontropic; neuroprotective;
osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
antidiabetic; hypotensive; dermatological; immunosuppressive;
antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
antithyroid; and antianemic. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
pathological conditions associated with an ORFX-associated disorder. The
nucleic acids can be used to express ORFX proteins in gene therapy
vectors. The proteins and nucleic acids may be used to treat cancers,
proliferative disorders, neurodegenerative disorders, osteoarthritis,
graft vs host disease, cardiovascular disease, diabetes mellitus,
hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
bacterial or fungal infection, malaria, autoimmune disorders, asthma,
allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
nocturnal haemoglobinuria, antinflammatory disease; to enhance
coagulation; to inhibit thrombosis; and as a contraceptive.
Sequence 324 BP; 82 A; 84 C; 89 G; 69 T; 0 other;
Query Match 85.5%; Score 18.8; DB 21; Length 324;
Best Local Similarity 90.9%; Pred. NO. 2.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ctgccagccggaagcctggg 22
|||||||
Db 71 CTGCCAGCGCGAGTGGG 50
RESULT 4
AAZ27241/c
ID AAZ27241 standard; cDNA; 1592 BP.
AC AAZ27241;
XX
XX 23-NOV-1999 (first entry)
XX Human secreted protein cDNA encoding gene 9.
XX Secreted protein; human; gene therapy; diagnosis; treatment; cancer;
KW protein therapy; tumor; neurodegenerative disorder; blood disorder; AIDS;
KW developmental abnormality; leukemia; immune system; autoimmune disease;
KW hepatic disease; renal disease; inflammation; allergy; schizophrenia;
KW Alzheimer's disease; cognitive disorder; arthritis; infection; psoriasis;
KW transplant rejection; diabetes; asthma; sepsis; acute; metabolic disorder;

KW cardiovascular disorder; food additive; preservative; ss.
XX Homo sapiens.
XX WO9946289-A1.
XX 16-SEP-1999.
XX 11-MAR-1999; 99WO-US05721.
XX 12-MAR-1998; 98US-0077686.
XX 12-MAR-1998; 98US-0077687.
XX 12-MAR-1998; 98US-0077696.
XX 12-MAR-1998; 98US-0077714.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Ferrie AM, Rosen CA, Florence C, Young PE, Yu G, NI J;
PI WPI; 1999-551363/46.
XX P-PSDB; AAY45265.
XX New isolated human genes, useful for diagnosis and treatment of, e.g.
XX cancers -
XX Claim la; 155; 306pp; English.
XX This invention describes novel human genes and the secreted proteins
XX they encode. The polynucleotides and their corresponding secreted
XX polypeptides are useful for preventing, treating or ameliorating medical
XX conditions e.g. by protein or gene therapy. Also pathological conditions
XX can be diagnosed by determining the amount of the new polypeptides in a
XX sample or by determining the presence of mutations in the new
XX polynucleotides. Specific uses are described for each of the
XX polynucleotides of the invention, based on which tissues they are most
XX highly expressed in, and include developing products for the diagnosis or
XX treatment of cancer, tumors, neurodegenerative disorders, developmental
XX abnormalities, blood disorders, leukemias, diseases of the immune system,
XX Alzheimer's and cognitive disorders, renal disease, inflammation, allergies,
XX infections, AIDS, transplant rejection, schizophrenia, arthritis,
XX psoriasis, cardiovascular disorders, diabetes, asthma, sepsis, acne,
XX polypeptides or polynucleotides can also be used as food additives or
XX preservatives. The polypeptides are also useful for identifying their
XX binding partners. AAZ27233-227265 encode human secreted proteins
XX described in the method of the invention.
XX Sequence 1592 BP; 432 A; 382 C; 418 G; 360 T; 0 other;
Query Match 85.5%; Score 18.8; DB 20; Length 1592;
Best Local Similarity 90.9%; Pred. NO. 2.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ctgccagccggaagcctggg 22
|||||||
Db 1066 CTGCCAGCGCGAGCTGAG 1045
RESULT 5
AAL34910/c
ID AAL34910 standard; cDNA; 313 BP.
XX AAL34910;
XX
XX 08-JAN-2002 (first entry)
XX Human musculoskeletal system related polynucleotide SEQ ID NO 252.
XX Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;

Thu Jun 20 06:56:44 2002

us-09-462-955b-1_copy_655_676.rng

KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ss.

XX Homo sapiens.

XX WO200155367-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01338.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180828.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226688.

XX 23-AUG-2000; 2000US-0227182.

PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-451937/48.
 DR P-PSDB; ABB03328.
 XX
 PT Isolated polypeptide for treating, preventing and/ or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 XX
 PS Claim 1; SEQ ID NO 252; 781pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 313 BP; 53 A; 90 C; 106 G; 62 T; 2 other;

Query Match 80.9%; Score 17.8; DB 22; Length 313;
 Best Local Similarity 90.5%; Pred. No. 5.5e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccaggccgaagcctggg 22
 ||||| ||||| ||||| |||||
 Db 89 TGCCAGGAGAGAGCCAGGG 69

RESULT 6
 AAZ38123/C
 ID AAZ38123 standard; cDNA; 496 BP.
 AC AAZ38123;
 XX
 XX 08-FEB-2000 (first entry)
 DE Human FATP alternatively spliced sequence.
 DE
 DE Fatty acid transport protein; FATP; hFATP; cardiomyopathy; diabetes;
 KW long-chain fatty acid metabolism; obesity; human; ss.
 KW
 OS Homo sapiens.
 OS
 PN WO951740-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 02-APR-1999; 99WO-EP02295.
 XX
 PR 06-APR-1998; 98EP-0400823.
 XX
 PA (JANC) JANSSEN PHARM NV.
 (UNIW) UNIV WASHINGTON.

XX Martin G, Nemoto M, Deeb SS, Auwerx J;
 PI WPI; 1999-620202/53.
 XX
 DR New human fatty acid transport protein, hFATP, useful to screen for
 DR inhibitors or enhancers useful to regulate fatty acid metabolism -
 XX
 PS Claim 2; Fig 1A-D; 83pp; English.
 XX
 CC The invention provides a human fatty acid transport protein (hFATP).
 CC hFATP is believed to be involved in the modulation long-chain fatty acid
 CC metabolism; the protein and polynucleotides therefore enable production
 CC of compositions comprising a component regulating (inhibiting or
 CC enhancing) expression of the hFATP gene useful therapeutically to alter
 CC intracellular or blood levels of long chain fatty acids. Such compounds
 CC are especially useful to treat conditions associated with deficient
 CC regulation (e.g. may comprise an inhibitor to treat cardiomyopathies or
 CC diabetes or an enhancer to treat obesity. The polynucleotides are also
 CC useful to screen compounds for their effects on hFATP expression, e.g.
 CC by measuring mRNA transcription in cells/cell extracts (e.g. liver
 CC cells) contacted with the compound and comparing with that in non-
 CC contacted cells. Sequences AAZ38122-123 represent nucleotide sequence of
 CC clones containing 2 kb sequences obtained from a cDNA library of human
 CC adipose tissue. The present sequence represents an alternative spliced
 CC sequence found in the analyzed clones.
 XX
 SQ Sequence 496 BP; 56 A; 151 C; 202 G; 87 T; 0 other;

Query Match 80.9%; Score 17.8; DB 20; Length 496;
 Best Local Similarity 90.5%; Pred. No. 5.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccaggccgaagcctggg 22
 ||||| ||||| ||||| |||||
 Db 435 TGCCAGGAGAGAGCCAGGG 415

RESULT 7
 AAZ49239/C
 ID AAZ49239 standard; cDNA; 524 BP.
 AC AAZ49239;
 XX
 XX 07-MAR-2000 (first entry)
 DE Human hydrolase homologue HHH-6 cDNA fragment 1514414F6.
 DE
 DE Hydrolase; homologue; HHH-6; N-terminal asparagine amidohydrolase;
 KW vanin-I; glycosyl hydrolase; glucosyl hydrolase;
 KW N-acetylglucosamine 6-P deacetylase; diagnosis;
 KW treatment; prevention; expression; disorder; carbohydrate; metabolism;
 KW antagonist; reproductive disorder; cell proliferation; ds.
 XX
 OS Synthetic.
 OS Homo sapiens.
 OS
 PN WO951626-A2.
 XX
 PD 02-DEC-1999.
 XX
 PF 28-MAY-1999; 99WO-US12021.
 XX
 PR 29-MAY-1998; 98US-0087236.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Bandman O, Hillman JL, Yue H, Lal P, Corley NC, Guegler KJ;
 PI Patterson C, Baughn MR;
 XX
 DR WPI; 2000-062716/05.
 XX

PT New human hydrolase homologues, useful for treating or preventing a
PT carbohydrate metabolism disorder -
PS Disclosures; Page 77; 91pp; English.
XX
XX Sequences AAZ49235, AAZ49237, AAZ49239, AAZ49242 and AAZ49259 represent
CC human hydrolase homologue HHH-6 cDNA fragments. The full-length
CC cDNA sequence is given in AAZ49228. Nucleic acids encoding HHH-6
CC were initially identified in a uterine endometrium tissue cDNA
CC library. HHH-6 has homology to N-acetylglucosamine 6-P
CC deacetylase. The invention relates to human hydrolase
CC homologues HHH-1 to HHH-7 (AAZ58165-Y58171) which respectively
CC have homology to N-terminal asparagine amidohydrolase, vanin-1,
CC glycosyl hydrolases, glucosylase and N-acetylglucosamine 6-P
CC deacetylase. Such homologues are useful in methods for diagnosing,
CC treating or preventing disorders associated with expression of
CC hydrolases. The hydrolase homologues are useful for treating or
CC preventing a carbohydrate metabolism disorder. Antagonists of these
CC hydrolases can be used to treat or prevent a reproductive or cell
CC proliferation disorder.
XX
SQ Sequence 524 BP; 95 A; 160 C; 167 G; 90 T; 12 other;

Query Match 80.9%; Score 17.8; DB 21; Length 524;
Best Local Similarity 90.5%; Pred. No. 5.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcacagccgagcagcctggg 22
| | | | | | | | | | | | | | | | | | | | | |
DB 299 TGCCAGGCCCCAGGCCAGGG 279

RESULT 8
AAS54242
ID AAS54242 standard; DNA; 807 BP.
XX
AC AAS54242;
XX
DT 13-FEB-2002 (first entry)
XX
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #373.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR P-PSDB; AAU36383.
DR
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX

PS Claim 27; Seq ID No 7879; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 807 BP; 115 A; 290 C; 289 G; 113 T; 0 other;

Query Match 80.9%; Score 17.8; DB 23; Length 807;
Best Local Similarity 90.5%; Pred. No. 5.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccagccgagcagcctggg 21
| | | | | | | | | | | | | | | | | | | | | |
DB 302 ccgcccagccgagcagcctggg 322

RESULT 9
AAS06749/c
ID AAS06749 standard; cDNA; 834 BP.
XX
AC AAS06749;
XX
DT 12-SEP-2001 (first entry)
XX
DE Polynucleotide sequence encoding human protein kinase #49.
XX
XX Human; protein kinase; PTK; STK; cancer; cardiovascular disease;
KW metabolic disorder; immune related disease; neurological disorder;
KW neurodegenerative disorder; inflammatory disorder; infectious disease;
KW reproductive disorder; gene therapy; ss.
XX
OS Homo sapiens.
XX
XX WO200138503-A2.
XX
PN 31-MAY-2001.
XX
PD 22-NOV-2000; 2000WO-US32085.
XX
PR 24-NOV-1999; 99US-0167482.
XX
PA (SUGE-) SUGEN INC.
XX
XX Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;
PI Flanagan P, Clary D;
XX
XX WPI; 2001-343950/36.
DR P-PSDB; AAU03549.
XX
XX Nucleic acids encoding human kinase polypeptides, useful for preventing
PT diagnosing and/or treating e.g. cancer, immune, cardiovascular and
PT neuronal-associated diseases, and microbial infections -
XX
XX Example 1; Figure 1; 433pp; English.

CC AAS06701-AAS06757 encode for novel human protein kinases #1-57. The
 CC novel protein kinases have been identified as members of the tyrosine
 CC or serine/threonine kinase (PTK and STK) families. The polynucleotides
 CC encoding protein kinases and the polypeptides may be used in the
 CC prevention, diagnosis and treatment of diseases associated with
 CC inappropriate kinase expression. For example, they may be used to treat
 CC cancers (especially cancers of haematopoietic origin), cardiovascular
 CC disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),
 CC immune related diseases (e.g. rheumatoid arthritis), neurological
 CC disorders (e.g. schizophrenia), neurodegenerative disorders (e.g.
 CC Parkinson's disease), inflammatory disorders (e.g. asthma), infectious
 CC disease (e.g. HIV) and reproductive disorders (e.g. infertility).
 CC Additionally, polynucleotides encoding protein kinases may be
 CC used for gene therapy and as DNA probes in diagnostic assays.
 CC The protein kinase polypeptides may be used as antigens in the production
 CC of antibodies against the protein kinases and in assays to identify
 CC modulators of protein kinase expression and activity.
 XX
 SQ Sequence 834 BP; 229 A; 177 C; 209 G; 219 T; 0 other;

Query Match 80.9%; Score 17.8; DB 22; Length 834;
 Best Local Similarity 90.5%; Pred. No. 5,1e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccagggcgaaggcctgg 21
 |||||
 Db 493 CTGCCAGGCCAAGCGCTGG 473

RESULT 10
 AAL01852/c
 ID AAL01852 standard; cDNA; 1092 BP.
 XX
 AC AAL01852;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE Human reproductive system related antigen cDNA SEQ ID NO: 1853.
 XX
 KW Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200155320-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01339.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216847.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220563.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 23-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0232081.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0241826.
 PR 08-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.

Thu Jun 20 06:56:44 2002

08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249219.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-465570/50.
P-PSDB; AAM95882.
Isolated nucleic acid molecule encoding a reproductive system antigen
is used in preventing, treating or ameliorating a medical condition -
Claim 1; SEQ ID NO 1853; 1297pp + Sequence Listing; English.
The present invention provides the protein and coding sequences of a
number of human reproductive system related antigens. These can be used
in the prevention and treatment of reproductive system disorders,
including cancer. The present sequence is a coding sequence of the
invention.
sequence 1092 BP; 280 A; 267 C; 280 G; 260 T; 5 other;
Query Match 80.9%; Score 17.8; DB 22; Length 1092;
Best Local Similarity 90.5%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ctgccagggccgaagcctgg 21
|||||
Db 685 CTGCCAGGCCAACAGCGCTGG 665
RESULT 11

AAS27197/c
ID AAS27197 standard; cDNA; 1092 BP.
XX
AC
AAS27197;
XX
DT
07-NOV-2001 (first entry)
XX
DE
cDNA encoding novel signal transduction pathway protein, Seq ID 232.
XX
KW
Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
immune system disorder; rheumatoid arthritis; inflammatory condition;
organ transplant rejection; infection; hepatitis C; blood disorder;
sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
reproductive system; gastrointestinal; liver disorder; AIDS; ss;
acquired immune deficiency syndrome.
XX
OS
Homo sapiens.
XX
PN
WO200154733-A1.
XX
PD
02-AUG-2001.
XX
PF
17-JAN-2001; 2001WO-US01312.
XX
PR
31-JAN-2000; 2000US-0179065.
PR
04-FEB-2000; 2000US-0180828.
PR
24-FEB-2000; 2000US-0184664.
PR
02-MAR-2000; 2000US-0186350.
PR
16-MAR-2000; 2000US-0189874.
PR
17-MAR-2000; 2000US-0190076.
PR
18-APR-2000; 2000US-0198123.
PR
19-MAY-2000; 2000US-0205515.
PR
07-JUN-2000; 2000US-0209467.
PR
28-JUN-2000; 2000US-0214886.
PR
30-JUN-2000; 2000US-0215135.
PR
07-JUL-2000; 2000US-0216647.
PR
07-JUL-2000; 2000US-0216880.
PR
11-JUL-2000; 2000US-0217487.
PR
11-JUL-2000; 2000US-0217496.
PR
14-JUL-2000; 2000US-0218290.
PR
26-JUL-2000; 2000US-0220963.
PR
26-JUL-2000; 2000US-0220964.
PR
14-AUG-2000; 2000US-0224519.
PR
14-AUG-2000; 2000US-0225213.
PR
14-AUG-2000; 2000US-0225214.
PR
14-AUG-2000; 2000US-0225266.
PR
14-AUG-2000; 2000US-0225267.
PR
14-AUG-2000; 2000US-0225268.
PR
14-AUG-2000; 2000US-0225270.
PR
14-AUG-2000; 2000US-0225447.
PR
14-AUG-2000; 2000US-0225757.
PR
14-AUG-2000; 2000US-0225758.
PR
14-AUG-2000; 2000US-0225759.
PR
18-AUG-2000; 2000US-0226279.
PR
22-AUG-2000; 2000US-0226681.
PR
22-AUG-2000; 2000US-0226868.
PR
22-AUG-2000; 2000US-0227182.
PR
23-AUG-2000; 2000US-0227009.
PR
30-AUG-2000; 2000US-0228924.
PR
01-SEP-2000; 2000US-0229287.
PR
01-SEP-2000; 2000US-0229343.
PR
01-SEP-2000; 2000US-0229344.
PR
01-SEP-2000; 2000US-0229345.
PR
05-SEP-2000; 2000US-0229509.
PR
05-SEP-2000; 2000US-0229513.
PR
06-SEP-2000; 2000US-0230437.
PR
06-SEP-2000; 2000US-0230438.
PR
08-SEP-2000; 2000US-0231242.


```

PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251030.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(PHUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
P-PSDB; AAU17280.
WPI; 2001-465460/50.
Novel polypeptides useful for diagnosing, treating, preventing and/or
prognosing disorders related to the proteins, including cancers, immune
disorders and neuronal disorders -
Claim 1; SEQ ID NO 232; 880pp; English.
The invention relates to novel isolated polypeptides (I), and
polynucleotides (II). (I), (II) and the antibody to (I) are useful for
diagnosing, preventing and treating diseases including immune system
disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
transplant rejections and graft versus host disease, infectious diseases
(e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
other blood-related disorders (sickle cell anaemia), myeloproliferative
disorders, primary haematopoietic disorders, hyperproliferative
disorders (e.g. Gaucher's disease and cancer), neurodegenerative
disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
disorders (e.g. glomerulonephritis), cardiovascular disorders
(e.g. arrhythmia), respiratory disorders, dermatological disorders, in
wound healing, epithelial cell proliferation, endocrine disorders (e.g.
Addison's disease), reproductive system disorders, gastrointestinal
disorder (inflammatory disorders), liver disorders (cirrhosis),
as stimulators of B-cell responsiveness to pathogens, activators of
T-cells, to induce higher affinity antibodies, and as a means to induce
tumour proliferation in pathologies e.g. acquired immune deficiency
syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX

Query Match 80.9%; Score 17.8; DB 22; Length 1092;
Best Local Similarity 90.5%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccagggccgaagcctgg 21
| | | | | | | | | | | | | | | | | | | | |
Db 685 CTGCCAGGCCCAACGCCTGG 665

RESULT 12
AAFI8209/c
ID AAFI8209 standard; DNA; 1658 BP.
XX
AC AAFI8209;
XX
DT 14-MAR-2001 (first entry)
XX

```

DE Lung cancer associated polynucleotide sequence SEQ ID 228.
 XX Human; lung cancer associated protein; neuroprotective; cytostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnerary;
 KW gastrointestinal; nephrotropic; antifungal; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200055180-A2.
 XX
 XX 21-SEP-2000.
 XX
 XX 08-MAR-2000; 2000WO-US05918.
 XX
 XX 12-MAR-1999; 99US-0124270.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 XX Ruben SM;
 XX
 XX WPI; 2000-587514/55.
 DR P-PSDB; AAB58333.
 XX
 XX Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 XX Claim 1; Page 689-690; 1425pp; English.
 XX
 XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
 CC general; nephrotropic; antifungal; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 XX Sequence 1658 BP; 386 A; 429 C; 374 G; 463 T; 6 other;
 SQ
 Query Match 80.9%; Score 17.8; DB 21; Length 1658;
 Best Local Similarity 90.5%; Pred. No. 4.9e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ctgccagccggaagcctgg 21
 Db 674 CTGCCAGGTAGAGGCGCTGG 654
 RESULT 13
 AAF89013/C
 ID AAF89013 standard; DNA; 1938 BP.
 XX
 XX AAF89013;
 AC
 XX 06-JUL-2001 (first entry)
 DT
 XX Murine FATP1 coding sequence SEQ ID NO: 29.
 DE
 XX

KW Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
 KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
 XX weight control; tuberculosis; TB; anti-fungal; ds.
 OS Mus musculus.
 XX
 XX WO200121795-A2.
 XX
 XX 29-MAR-2001.
 XX
 XX 21-SEP-2000; 2000WO-US25891.
 XX
 XX 23-SEP-1999; 99US-0405504.
 PR 23-SEP-1999; 99US-0405505.
 PR 16-DEC-1999; 99US-0465280.
 PR 17-FEB-2000; 2000US-0506252.
 PR 06-JUL-2000; 2000US-0611197.
 XX
 XX (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 XX Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;
 PI
 XX
 XX WPI; 2001-354783/37.
 DR
 XX
 XX New fatty acid transport proteins (FATPs) useful for the manufacture of
 PT medicament for treating obesity, diabetes and heart disease -
 PT
 XX
 XX Disclosure; Fig 30; 287pp; English.
 PS
 XX
 XX The present invention provides the protein and coding sequences of fatty
 CC acid transport proteins (FATPs) from a number of species, including
 CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
 CC from the mouse, FATP4 and b from C. elegans, and FATP from Aspergillus
 CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
 CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
 CC tuberculosis can be used to identify inhibitors which can then be used to
 CC treat TB. That from M. grisea (also known as rice blast fungus) can be
 CC used to develop anti-fungal agents capable of preventing infection of
 CC rice. Those from the human can be used to develop treatments for
 CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
 CC present sequence is one of the sequences described in the exemplification
 CC of the invention.
 XX
 XX Sequence 1938 BP; 349 A; 571 C; 615 G; 403 T; 0 other;
 SQ
 Query Match 80.9%; Score 17.8; DB 22; Length 1938;
 Best Local Similarity 90.5%; Pred. No. 4.8e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tgcccagccggaagcctggg 22
 Db 517 TGCCCAGGCAGAGGCCAGGG 497
 RESULT 14
 AAF89012/C
 ID AAF89012 standard; DNA; 1941 BP.
 XX
 XX AAF89012;
 AC
 XX 06-JUL-2001 (first entry)
 DT
 XX Human FATP1 coding sequence SEQ ID NO: 28.
 DE
 XX
 XX Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
 KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
 KW weight control; tuberculosis; TB; anti-fungal; ds.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200121795-A2.
 PN

XX PD 29-MAR-2001.
 XX PF 21-SEP-2000; 2000WO-US25891.
 XX PR 23-SEP-1999; 99US-0405504.
 XX PR 23-SEP-1999; 99US-0405505.
 XX PR 16-DEC-1999; 99US-0455280.
 XX PR 17-FEB-2000; 2000US-0506252.
 XX PR 06-JUL-2000; 2000US-0611197.
 XX PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 XX PA (MILL-) MILLENNIUM PHARM INC.
 XX PI Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;
 XX WPI: 2001-354783/37.
 XX DR New fatty acid transport proteins (FATPs) useful for the manufacture of
 XX PT medicament for treating obesity, diabetes and heart disease -
 XX PS Disclosure; Fig 30; 287pp; English.
 XX CC The present invention provides the protein and coding sequences of fatty
 XX CC acid transport proteins (FATPs) from a number of species, including
 XX CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
 XX CC from the mouse, FATP4 and b from C. elegans, and FATP from Aspergillus
 XX CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
 XX CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
 XX CC treat TB, that from M. grisea (also known as rice blast fungus) can be
 XX CC used to develop anti-fungal agents capable of preventing infection of
 XX CC rice. Those from the human can be used to develop treatments for
 XX CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
 XX CC present sequence is one of the sequences described in the exemplification
 XX CC of the invention.
 XX SQ Sequence 1941 BP; 325 A; 620 C; 651 G; 345 T; 0 other;
 Query Match 80.9%; Score 17.8; DB 22; Length 1941;
 Best Local Similarity 90.5%; Pred. No. 4.8e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tgcccagggccgaagcctggg 22
 ||||||| ||||| |||
 Db 517 TGCCAGGAGGAGGCGG 497
 RESULT 15
 AAZ38122/c
 ID AAZ38122 standard; cDNA; 2219 BP.
 XX AC AAZ38122;
 XX DT 08-FEB-2000 (first entry)
 XX DE Human FATP protein encoding cDNA.
 XX KW Fatty acid transport protein; FATP; hFATP; cardiomyopathy; diabetes;
 XX KW long-chain fatty acid metabolism; obesity; human; ss.
 XX OS Homo sapiens.
 XX PN WO9951740-A2.
 XX PD 14-OCT-1999.
 XX PF 02-APR-1999; 99WO-EP02295.
 XX PR 06-APR-1998; 98EP-0400823.
 XX PA (JANC) JANSSEN PHARM NV.

PA (UNIW) UNIV WASHINGTON.
 XX PI Martin G, Nemoto M, Deeb SS, Auwerx J;
 XX DR WPI: 1999-620202/53.
 XX DR P-PSDB; AAY40435.
 XX PT New human fatty acid transport protein, hFATP, useful to screen for
 XX PT inhibitors or enhancers useful to regulate fatty acid metabolism -
 XX PS Claim 2; Fig 1A-D; 83pp; English.
 XX CC The invention provides a human fatty acid transport protein (hFATP).
 XX CC hFATP is believed to be involved in the modulation long-chain fatty acid
 XX CC metabolism; the protein and polynucleotides therefore enable production
 XX CC of compositions comprising a component regulating (inhibiting or
 XX CC enhancing) expression of the hFATP gene useful therapeutically to alter
 XX CC intracellular or blood levels of long chain fatty acids. Such compounds
 XX CC are especially useful to treat conditions associated with deficient
 XX CC regulation (e.g. may comprise an inhibitor to treat cardiomyopathies or
 XX CC diabetes or an enhancer to treat obesity. The polynucleotides are also
 XX CC useful to screen compounds for their effects on hFATP expression, e.g.
 XX CC by measuring mRNA transcription in cells/cell extracts (e.g. liver
 XX CC cells) contacted with the compound and comparing with that in non-
 XX CC contacted cells. Sequences AAZ38122-123 represent nucleotide sequence of
 XX CC clones containing 2 kb sequences obtained from a cDNA library of human
 XX CC adipose tissue. The present sequence represents a cDNA encoding the
 XX CC hFATP protein.
 XX SQ Sequence 2219 BP; 369 A; 710 C; 731 G; 406 T; 3 other;
 Query Match 80.9%; Score 17.8; DB 20; Length 2219;
 Best Local Similarity 90.5%; Pred. No. 4.8e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tgcccagggccgaagcctggg 22
 ||||||| ||||| |||
 Db 539 TGCCAGGAGGAGGCGG 519

Search completed: June 19, 2002, 16:50:44
 Job time: 17502 sec

Thu Jun 20 06:56:44 2002

us-09-462-955b-1_copy_655_676.rng